Quantitative Trait Locus Mapping Reveals Regions of the Maize Genome Controlling Root System Architecture

by

Paul Roman Zurek

University Program in Genetics and Genomics
Duke University

Date:

_____

Approved:

_____
Philip Benfey, Supervisor

_____
Thomas Mitchell-Olds

_____
Paul Magwene

_____
Meng Chen

Dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy in the University Program in
Genetics and Genomics
of Duke University

2014

ABSTRACT

Quantitative Trait Locus Mapping Reveals Regions of the Maize Genome Controlling Root System Architecture

by

Paul Roman Zurek

University Program in Genetics and Genomics
Duke University

Date:

_____

Approved:

_____
Philip Benfey, Supervisor

_____
Thomas Mitchell-Olds

_____
Paul Magwene

_____
Meng Chen

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the University Program in Genetics and Genomics in the Graduate School of Duke University

2014

Abstract

Root system architecture (RSA) is the spatial distribution of roots of individual plants. As part of a collaborative effort I adapted a gellan gum based system for imaging and phenotyping of root systems in maize. This system was first used to perform a survey of 26 distinct maize varieties of the Nested Association Mapping (NAM) population. The analysis of these data showed a large amount of variation between different RSA, in particular demonstrating tradeoffs between architectures favoring sparse, but far reaching, root networks versus those favoring small but dense root networks. To study this further I imaged and phenotyped the B73 (compact) x Ki3 (exploratory) mapping population. These data were used to map 102 quantitative trait loci (QTL). A large portion of these QTL had large, ranging from 5.48% to 23.8%. Majority of these QTLs were grouped into 9 clusters across the genome, with each cluster favoring either the compact of exploratory RSA. In summary, our study demonstrates the power of the gellan based system to locate loci controlling root system architecture of maize, by combining rapid and highly detailed imaging techniques with semi-automated computation phenotyping.

Dedication

To my significant other Josie
To my parents Bożena and Andrzej
To my brother Marcin.
Thank you.

Contents

List of Tables

List of Figures

Acknowledgements

Thank you to my advisor, Philip Benfey, for letting me work in his lab. It has been both a privilege and pleasure to learn from him. Thank you to all the wonderful people in the lab, past and present, for both their scientific input and social shenanigans. Thank you to my committee Tom, Meng and Paul for all their help and mentorship through my years at Duke. Special thank you to Chris Topp, for being my go to guy and postdoc mentor.

Chapter 1. Introduction

This thesis presents the work involved in developing of a gel-based system for capturing highly detailed images and phenotyping of root system architecture (RSA), which is the spatial distribution of roots of individual plants. It describes the use of that system to survey 26 distinct maize varieties, followed by imaging of a Recombinant Inbred Line (RIL) population and Quantitative Trait Loci (QTL) mapping. Chapter 1 consists of an introduction to relevant background information. Chapter 2 describes the development and testing of the system itself. Chapter 3 demonstrates the first use of the gel system to survey the 26 founder lines of the Nested Association Mapping (NAM) population. Chapter 4 includes both the phenotyping and QTL analysis of the B73 x Ki3 RIL population, and Chapter 5 summarizes and concludes this work.

Importance of root system architecture

Plants are absolutely necessary for the continual survival of humanity. They are a direct source of food, and indirectly as feed for livestock. The improvement of key crop plants will continue to increase in importance in order to keep up with ever growing human population. One such key crop, maize, has already seen an amazing amount of improvement since its humble beginnings as teosintes around 9000 years ago (Doebley et al., 1990; Wang et al., 2005; Matsuoka et al., 2002). Although the initial yield increases were due to simple breeding selection by early farmers, some of the more recent improvements

have been due to things such as the Green Revolution (Sasaki et al., 2002; Peng et al., 1999), genetically modification (making them GMOs) and use of inorganic fertilizers, herbicides and pesticides. As a result, the recorded yield measurements over the last 200 years show more than a 10-fold yield improvement (Duvick, 2005). Maize is especially amenable to improvements because of the great amount of genetic diversity that exists in this species (Flint-Garcia et al., 2005; Wright et al., 2005). It's often said that there can be as much variation between any two cultivars of maize as there is between humans and monkeys.

This variability has been put to good use to improve maize yield, mostly by focusing on above-ground parts of the plant. Traits such as number of kernels per ear or numbers of ears, steeper leaf angles to allow higher plant density (Hammer et al., 2009), increased resistance to both biologically (Balint-Kurti and Carson, 2006) or chemically (Magalhaes et al., 2007) induced stress have all been ameliorated. Until recently, few studies have been focused on root system architectures (RSA) as a means of to increasing plant yield. This is unfortunate as RSA can potentially have a great effect on plant health and yield, and selecting varieties with the appropriate RSA for different growth conditions could have highly beneficial conditions. Several studies have already reported advantages of certain types of root architectures for different conditions. In maize, steeper root angles have been shown to allow for higher planting densities by decreasing the root footprint allowing more plants to be planted in the same area without negative effects

(Hammer et al., 2009). Deep roots have been shown to have beneficial effects in several species under drought conditions (Ho et al., 2005; Lynch, 2013). Most recently the DEEPER ROOTING 1 (DRO1) rice quantitative trait locus, which results in longer rice roots, has been shown to increase yield in rice under drought conditions due to its ability to reach deep subsurface water (Uga et al., 2013). On the other hand, plants with roots closers to the top soil are better at acquiring nutrients, such as phosphorus, which tend to be found there (Jobbágy and Jackson, 2001). This has been shown in rice, where the gene Pstol1 resulted in increased yields in phosphorus poor soils as a result of increased root mass (Gamuyao et al., 2012). Despite this, little work has been done to locate genes responsible for root architecture of crop plants. This is in large part due to difficulties associated with phenotyping root systems.

Root phenotyping

The main reason for the disproportionate amount of aerial versus below ground studies is simply the combination of ease of phenotyping aerial tissues and the difficulties of phenotyping root systems. The spatial organization of roots, or root system architecture (RSA) of monocot plants like corn can be especially challenging due to their complexity and size (Hochholdinger et al., 2004). A number of approaches have been utilized to explore RSA. These range in their throughput rates, level of phenotyping detail, degree of

automation, technical expertise required and cost. All excel in some areas, but demonstrate weaknesses in others.

The most straightforward of these is "shovelomics" (Trachsel et al., 2010), which involves digging out the root crown from soil, washing it, and measuring several traits using a standardized scoreboard. This scoreboard includes several rulers and protractors that are used to measure root lengths and angles. This is done with both the whole excavated crown as well as sample roots taken from the crown. The advantage of this approach is that it allows collection of a large number of samples in a relatively small amount of time. Additionally, the phenotyping can be done with adult, field grown plants. Unfortunately this system also has weaknesses. Only a portion of the root system is excavated. Most of the smaller and outlying roots cannot be dug up. The measurements themselves are performed on the crown lying flat on the scoreboard which ignores the 3D RSA contained within the crown. Lastly, only a few of the smaller roots are measured and the resulting trait values are extrapolated to the whole root system. Although this method can be high throughput without need for expensive equipment or specialized expertise, it provides only a broad overview of RSA, and as such it is very good at phenotyping a large number of different lines at a coarse level.

Although using soil as growth media would be ideal, excavations of whole root systems from soil can be extremely time intensive (Weaver and Bruner, 1927).

Shovelomics approaches have decreased the excavation time by focusing on a small portion of the root system (Trachsel et al., 2010). One way to increase throughput is to move away from soil and use media which allow for more rapid isolation and phenotyping of roots.

One such approach involves the use of hydroponics (Tuberosa et al., 2002). The plants are grown in a nutrient solution and the whole root system can be phenotyped as needed. This also has the advantage of allowing repeated measurements. As long as the phenotyping is performed nondestructively, the plants can be placed back into the nutrient solution. The experimental conditions can be easily controlled as the nutrient solution can be tailored to specific experiments. Nutrient concentration can be varied and toxic substances, such as salt or heavy metals, can be added to study their effects. Hydroponics systems have their weaknesses as well. Because containers are needed to grow the plants, fewer plants can be grown at any time. Because the roots are free floating in the solution, their 3D architecture does not reflect the root distribution of plants grown in a solid media. Phenotyping itself is usually done by eye, either through direct measurements, or by scans of the root systems, which results in loss of detailed information.

A complementary system was developed by (Hund et al., 2009; Le Marié et al., 2014). The plants are grown sandwiched between two sheets of germination paper placed

in a plastic bag. The bags are placed upright in shallow containers filled with nutrient solution, allowing it to be wicked up throughout the germination paper. Once the plants reach the desired age, they're removed, the roots scanned and analyzed using WinRhizo or similar root phenotyping software. This systems allows phenotyping of a large number of plants while preserving some of the spatial distribution of the root systems. Unfortunately, unlike true hydroponic system, the small size of the pouches limits the paper pouch experiments to young plants.

All of the approaches described so far demonstrate a high level of throughput but a low level of detail. Several medical imaging technologies, such as x-ray micro-CT (Hargreaves et al., 2008) or combined PET-MRI (Jahnke et al., 2009), have been used to produce much more detailed phenotyping of RSA. These systems first use the relevant technology to image the root system, and then create a 3D representation that can be used for highly detailed phenotyping. The maximum size of the root is dependent on the size of the machine being used, but in most cases this is limited to younger plants. The plants can be also be grown in either a soil or nutrient media. Although this makes it very appealing, these systems have several weak points. They require expensive and dedicated equipment that might not be available to many researchers. This has the effect of severely reducing the imaging throughput as likely only a single machine would be available for use. The scans themselves can also be time-consuming. Additionally these types of

technologies are dependent on computational phenotyping, which requires much more expertise to setup and use than, for example, Shovelomics.

Although these types of technologies are not conducive to large scale experiments, they are much better for focused studies. A good example is the use of radiolabeled compounds and microPET (Ferrieri et al., 2013). Short lived radioactively labeled compounds can be used to trace their incorporation and distribution throughout the plants. For example radioactively labeled CO^2 from shoots can be tracked as it is distributed and allocated through the plant.

Gellan gum based imaging system

From the description above, it should be clear that no ideal root phenotyping system exists. On one end of the spectrum are approaches that can be used to coarsely phenotype a large number of plants, while on the other end are technologies that can give a very clear picture of the RSA, but are limited in throughput.

We designed and built our own phenotyping system, which resides somewhere between the above discussed approaches. This system uses a solidified hydroponic media. The plants are grown in glass containers and imaged on a dedicated imaging platform. Because of the translucent nature of the media and glass, the root system can easily be photographed. Due to this nondestructive approach, each plant cylinder can be imaged multiple times without disturbing the plants. The imaging itself is performed on a

turntable, and each plant is imaged multiple times from different angles, allowing us to fully capture the 3D architecture of the root system. This produced highly detailed images of the whole root system (Fig 1). Concurrently with the development of the imaging setup, we have, in collaboration with (Galkovskyi et al., 2012), developed the software necessary to process and analyze the image sets. The development and details of this system are further discussed in Chapter 2.

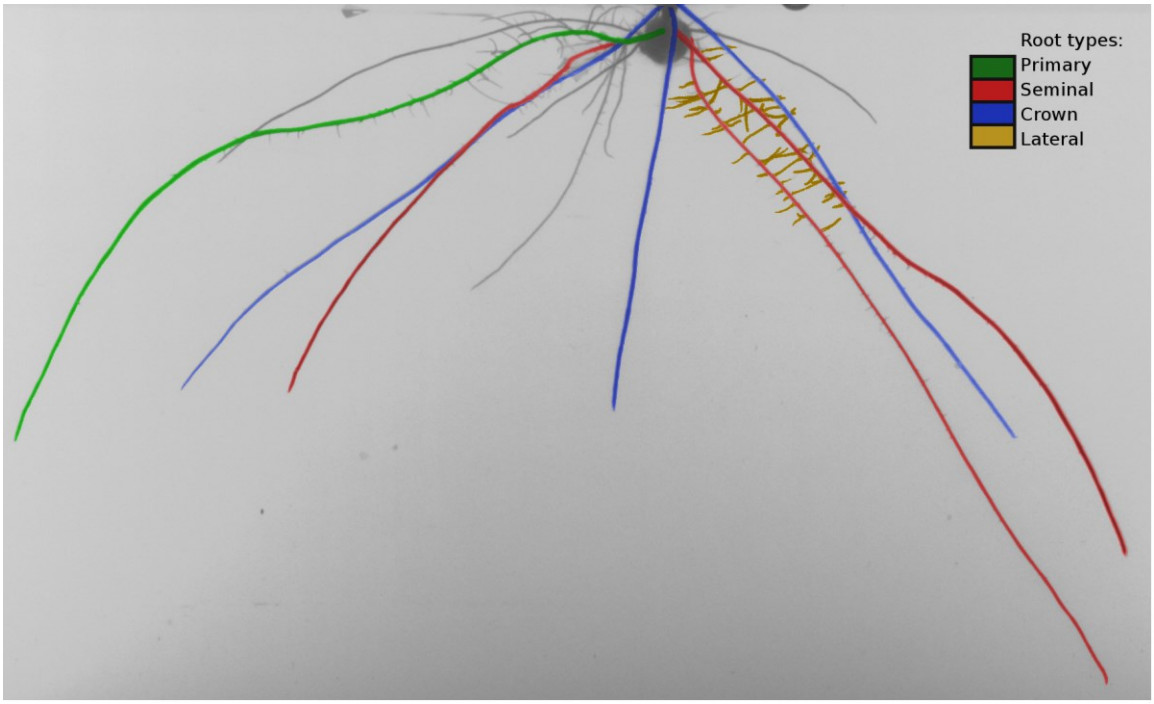


Figure 1: Example of an image from the gellan gum based imaging system, with manually annotated root types.

Nested Association Mapping (NAM) population

The nested association mapping (NAM) population was selected as the source of germplasm for the experiments presented in Chapters 3 and 4. The 26 founders of the

population were used to conduct a survey of maize RSA, while one of the subpopulations, was used for root trait QTL mapping. The NAM population was created as a collaboration of several maize labs (Yu et al., 2008; Buckler et al., 2009; McMullen et al., 2009). The population is composed of 25 different RIL families, each of which share a single common parent. Each subpopulation contains 200 individual lines, for a total of 5000 lines. The B73 variety was chosen as the common parent due to its history as an agronomically important line as well as its choice for the maize whole genome sequencing project (Schnable et al., 2009). The remaining varieties were selected using a computational approach to maximize the combined genetic diversity represented by all of the founders (Yu et al., 2008). This was based on 94 microsatellite markers, and was later confirmed using SNP markers.

The original goal for the NAM population was to map complex genetic traits (Yu et al., 2008). Mapping of such traits is generally accomplished through either linkage analysis or association mapping (Buckler et al., 2009). The former approach uses a manmade mapping population to find linkages between phenotypes and genotypes. Historically this has produced poor results as the resolution of QTL mapping is dependent on the number of recombination events in the mapping population as well as the number of genetic markers. Since the mapping populations used in these types of experiments are created in the lab, they contain relatively few recombination events and therefore the mapping resolution tends to be poor. On the other hand, association mapping uses large

panels of different varieties of the same species. This gives it many more recombination events as it takes advantage of the historical recombination in the species. This also requires many more markers, preferably at least one marker for each genetic loci. Only recent advances in high-throughput genomics, such as SNP chips and next-generation sequencing, have made this type of mapping possible. The NAM population is a combination of both approaches. It uses both the historical differences between the 26 founder lines, but also introduces man-made recombinations. The end result is that fewer markers are required than would be needed for a pure association study, but at the same time the resolution is much higher than for a pure QTL study (Pearson and Manolio, 2008). Here we focused on phenotyping one of the NAM subpopulations to perform a QTL mapping study, as it would not be feasible to phenotype the whole population in any reasonable time frame.

Quantitative Trait Locus (QTL) mapping

To locate genes that control root system architecture (RSA), we've taken a Quantitative Trait Locus (QTL) mapping approach. QTL mapping works by taking advantage of man-made recombination events between two phenotypically different founder varieties (Mauricio, 2001). Two varieties are first crossed and then a number of lines selected for inbreeding through 6 or more generations of single seed descent. The resulting Recombinant Inbred Line (RIL) population is composed of individuals with a

mosaic genome from the original founders (Fig 2). This population is phenotyped for traits of interest, as well as genotyped to determine recombination breakpoints. The combination of the two allows the mapping of specific loci linked to the measured phenotypes (Zeng, 1994; Basten et al., 1994).

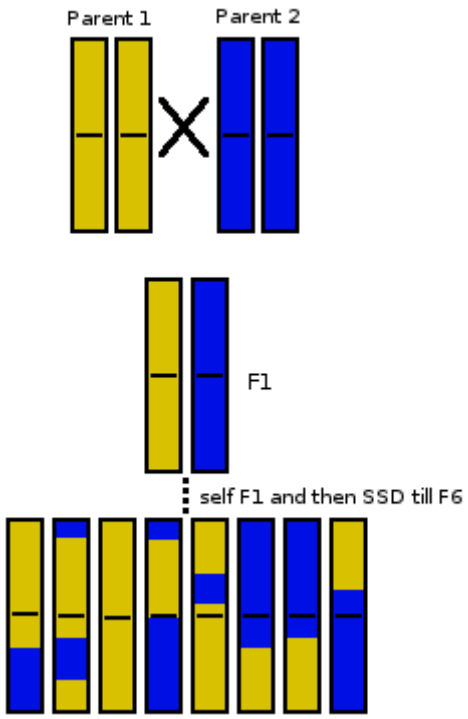


Figure 2: Diagrammatic representation of generation of a RIL mapping population.

Chapter 2. Development of the gel-based imaging system

As described in Chapter 1, work on the root systems of large crop plants, such as rice or maize, has recently begun to garner more interest. This is, in great part, due to development of new technologies, as well as application of older ones to the problem of exploring the below ground aspects of plants. This chapter is devoted to the steps taken by me with the help of other lab members in adaptation the gel system previously used for rice to maize experiments (Iyer-Pascuzzi et al., 2010; Clark et al., 2011; Fang et al., 2013). Although the rice protocols were a good entry point, most had to be significantly altered to be made suitable for corn. The resolutions of these corn specific problems, as well as more general challenges, have driven an overall improvement of the gel imaging system for use in rice and corn experiments.

Germination of maize for gel system experiments

The first challenge was growing maize in the gel system. There were several unsuccessful attempts by other lab members to use the same methods as for rice. At the time the rice seeds were surface sterilized and then planted directly in the gel about a centimeter under the surface. The same approach with corn resulted in an extremely low germination rate - less than 1 in 20 plants germinated. Those seed that did germinate produced stunted and unhealthy looking plants. The root systems were small and grew very slowly, while the shoots were brown and wilted. In many cases the coleoptile would

not split, and so the shoot would try to force itself out and end up damaging itself in the process (Abendroth et al., 2011). Several experiments with slight variations in the protocol, such as using different media or different surface sterilization protocols did not improve the germination rate.

Preliminary experiments did show that seeds planted close to the surface could grow. Unfortunately, roots that emerged from these seeds tended to grow on the surface of the gel. The surface tension of the gel would prevent the roots from penetrating into it, resulting in the roots growing on top of the gel surface. Only once a root reached the edge of a jar would it be able to start growing down into the gel.

Although this was far from ideal, it provided a clue as to what was happening. One of the key differences between rice and maize is that rice is an aquatic plant, it can germinate and grow while flooded. On the other hand, corn is a terrestrial plant and flooding tends to inhibit its germination (Martin et al., 1991; Armstrong et al., 1994). Once past the germination stage in a flooded environment, corn is able to grow in hydroponics-type systems as evidenced by many studies of that type (Holloway et al., 2011; Li et al., 2011). The germination issues I was having likely stemmed from this difference between rice and maize. Although the gel system is solid, most of the media is aqueous. The challenge was how to germinate corn in the gel system, since placing the seed within the gel inhibited germination, while placing the seed on top resulted in roots growing on the

surface of the gel. The solution was to first pre-germinate the seeds outside the growth container and then transplant them to gel after emergence of the primary root and coleoptile. This has proven an effective approach, and I have developed a detailed protocol for it. The seeds were first surface sterilized, then individually germinated in small petri dishes in a small amount of water, and lastly the seedlings were transplanted to jars and imaged once they reached the appropriate age.

Seed surface sterilization

Prior to germination, the seeds were first surface sterilized to eliminate fungal spores and bacteria. These contaminations not only occluded the root system, but probably also had effect on the root system architecture itself (Lugtenberg and Kamilova, 2009; Suslow and Schroth, 1982). I began by using the approach for surface sterilization of rice seeds (Iyer-Pascuzzi et al., 2010). This protocol involved a dilute bleach wash, followed by an ethanol wash and lastly several water washes. This proved ineffective for corn. The majority of seedlings sterilized in this way still ended up with fungal contamination after only a few days in the growth container. A more robust sterilization protocol was required. After some trial and error, a protocol using two hydrogen peroxide steps separated by several hours of imbibition in water has given much better results. The goal of water imbibition between the hydrogen peroxide treatments was to soften the hull of the maize kernel and allow the second hydrogen peroxide treatment to penetrate

further into the kernel. Although this protocol did not eliminate contamination entirely, it minimized it to a point where experiments became feasible, with an average of only 10-20% of plants becoming contaminated through their growth period. The sterilization protocols were slightly different for those used in the experiments in Chapter 3 versus those in Chapter 4.

The protocol for Chapter 3 was as follows:

1. The seeds were placed in 35% hydrogen peroxide on a shaker for 20 minutes. A modified monthly pill organizer was used for the sterilization (Fig 3).

2. The seeds were washed 3 times with sterile ddH20.

3. The seeds were transferred to a 24 well microtiter plate and incubated overnight at 30ºC in dark. The seeds were placed in a small amount of water (~15mL) such that each seed was in contact with the liquid, but halfway submerged by it.

4. The next morning the seeds were sterilized again with 35% hydrogen peroxide on a shaker for 10 minutes using the modified pill organizer.

5. The seeds were washed 3 times with sterile ddH20.

6. The seeds were placed in individual petri dishes [60mm x 15mm] with a small amount of water (~15mL) overnight at 30ºC in dark.

7. The seeds were planted on one of the two following days. Only seedling with an emergent coleoptile and ~1 inch long primary root were planted.

The microtiter plate and individual petri dishes were used to prevent contamination from spreading from seed to seed, as the protocol was never 100% effective and a single seed could contaminate the whole batch. Additionally the pill organizer and the microtiter plate were used to allow me to keep track of the weight of individual seeds throughout the protocol.

The protocol for Chapter 4 was as follows:

1. The seeds were placed in 35% hydrogen peroxide on a shaker for 20 minutes. A custom made sterilization chamber was used (Fig 3)

2. The seeds were washed once with sterile ddH20.

3. The seeds were left in the chamber, but enough water was added such that the seed were in contact with the water (~15mL), but not completely submerged by it.

4. The seeds were incubated at 30ºC in dark for 8 hours.

5. The seeds were sterilized with 35% hydrogen peroxide on a shaker for 20 minutes.

6. The seesd were placed in individual petri dishes [60mm x 15mm] with a small amount of water overnight (~15mL) at 30ºC in dark.

7. The seeds were planted on one of the two following days. Only seedling with an emergent coleoptile and ~1 inch long primary root were planted.

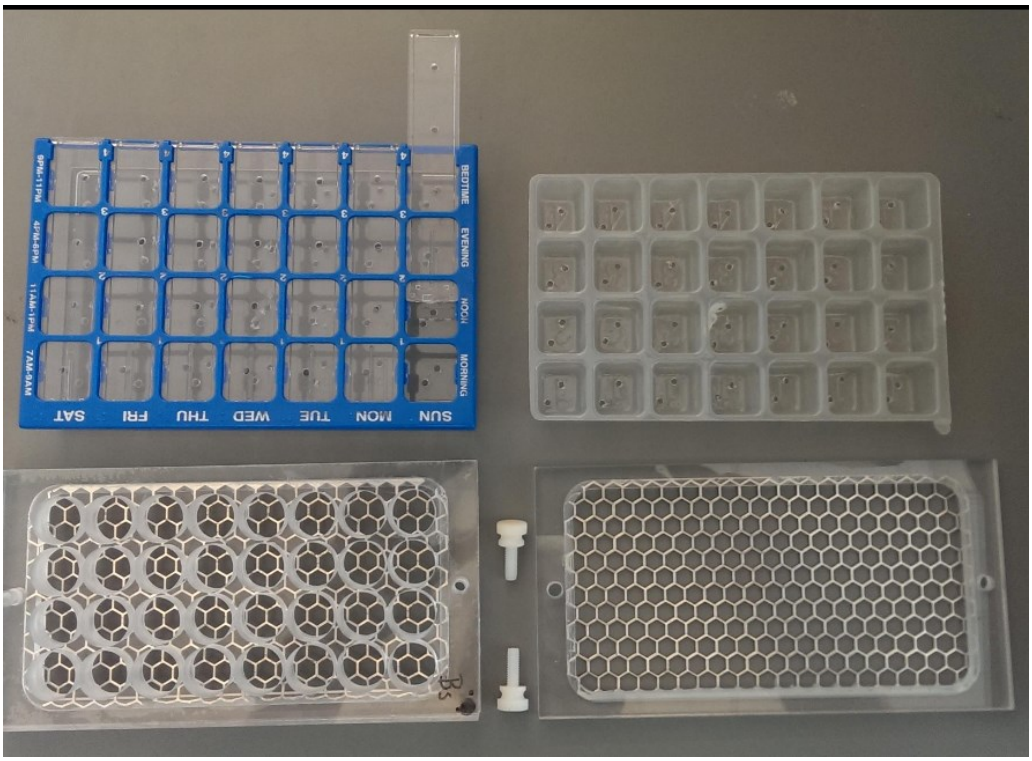


Figure 3: Photo of surface sterilization chambers.

Seedling Transplantation

The germination rate of the maize seeds was quite low, as on average only about half the seeds germinated. Furthermore, germination was asynchronous and could take 2, 3 or more days from the start of the sterilization. Ideally I would have liked to only plant seeds that germinated on the same day, but because of the wide range of times to germination, I used seeds that took 3 or 4 days to germinate. Whether the seedling used was 3 or 4 days old, I selected seedlings that were in the same early stages of growth, so even though planted on different days, the seedlings from day 3 and day 4 looked the same. These would have an emerging primary root about an inch long, and a visible

unopened coleoptile. Seedlings with large seminal roots were avoided if possible. I selected these to try to developmentally match the seed as closely as possible. To avoid confusion of different planting days, all references to imaging day assume day 0 to be the planting day and day 1 to be the first day after planting.

The planting itself was done as follows. The aluminum foil at the top of the jar was cut back, leaving only about half an inch from the edge. This was done to ensure that each jar would receive equal amounts of light. The amount of aluminum on the unplanted jars could vary greatly, and if left uncut the amount of light received by some plants would be far less than others. Additionally removing as much excess aluminum as possible maximized the amount of light that could enter the jar. In order to minimize contamination, 70% ethanol was liberally applied to the jars throughout the planting procedure.

The aluminum cap was then temporarily removed and a Pipe-Aid with a 25mL pipette was used to excavate a small, half-inch deep, hole in the middle of the jar. A pair of large flame-sterilized tweezers was then used to transfer a seedling to the hole. The seedling was placed such that the primary root was in the gel, while the coleoptile was above the surface of the gel. This was done by either sliding the primary root into the sidewall of the hole, or by using the tweezers to gouge a hole where the root would go. Damaging the gel matrix in such a way did not produce any artifacts during imaging, as

any depression in the gel would later fill with water from the surrounding gel matrix and wouldn't be noticeable in the images.

The most important step during planting was to place the seeding so that the coleoptile was above the surface of the gel. As long as that was ensured, the plant would grow. This was done because due to the large amount of water in the gel matrix the hole would rapidly fill and flood the seedling, inhibiting further growth. If at least a portion of the seedling was above the surface of the water, the transplanted seedling would readily grow. Additionally, during planting, the surface of the gel around the hole could be lightly scored with the tweezers. This was done so that if any of the roots that emerged from the seed ended up growing on the surface of the gel, they would have some place to latch onto and start growing into the gel. Lastly the aluminum foil cap was replaced on the jar, and sealed using micropore tape.

For the experiments in Chapter 3, the jars were directly moved to the growth chambers. All of the plantings were done late at night, so that he jars were moved to the growth chambers during their night cycle. As mentioned above, digging the hole during the planting, or even scoring the surface of the gel did not introduce any visual artifacts during imaging. This is because under high temperature and high lights in the growth chambers (28°C, 800 μmol) the gel underwent syneresis, resulting in some of the liquid in the gel matrix being released into the surrounding area. This resulted in flooding of any

uneven surfaces, eliminating any imperfections in the gel. Consequently, in order to give the newly planted seedling some time to grow before being flooded, the plants were transferred to the chambers while the temperature was lower (25°C) and the lights were off. For the experiments in Chapter 4, the jars were left overnight in a dark room at ambient temperature (20-21°C), and transferred to the growth chambers the following morning.

Hoagland's growth media

All of the earliest maize experiments were done using 1x Yoshida's, a media tailored for growing rice (Yoshida et al., 1976). Originally I suspected that the media might have been the cause of poor germination and decided to use a media more appropriate to maize. A literature search revealed that most of hydroponically grown maize was grown in one of several versions of Hoagland's media. Since the gel system is at its core a modified hydroponic system, I switched to using half-strength Hoagland's solution (Hoagland and Arnon, 1950) which is used at the Duke University Phytotron and Greenhouses (Table 1). This media has worked quite well as it provides an excess of nutrients needed by the plants, is easier to prepare than Yoshida's, and most importantly, the plants appear very healthy.

There was one minor problem with the Hoagland's media which was easily resolved. Several days after a jar had been transferred to the growth chamber, a brown

precipitate would appear. This precipitate would form as a diffuse band about one inch wide and two inches deep within the gel, but only at the interface of the gel and glass. Because of the color, I suspected that it was some kind of iron precipitate, either iron phosphate or iron oxide. Because the band only appeared at the interface of gel and jar and only closer towards the surface, which would be plentiful in oxygen, it was likely iron oxide, or rust. In either case, reducing the amount of iron in the solution by half prevented this band from appearing. This was necessary as this band would interfere with image processing, specifically image thresholding. Fortunately reducing the amount of iron did not have any impact on the health of the plants. Additionally because the maize plants are only grown for 12 days in the gel system, it's likely that a large portion of that growth was being supported by nutrient stores in the maize kernel alone.

	Solution A (for 1L sol'n)
Ca(NO3)2*4H2O	298.246 g
NH4NO3	20.569 g
	Solution B (for 1L sol'n)
KNO3	152.02 g
MgSO4*7H2O	126.65 g
ZnSO4*7H2O	0.286 g
MnSO4*H2O	0.195 g
CuSO4*5H2O	0.021 g
H3BO3	0.725 g
Na2MoO3*2H2O	0.011 g
Kcl	3.154 g
NH4H2PO4	59.03 g
	FeDPTA (for 1L sol'n)
FeDPTA (Sequestrene 330)	37.5 g
All solutions are fileter sterilized. FeDPTA is stored in dark.	

Table 1: Recipe for the Hoagland's media used for all maize experiments.

Maize growth containers

When it became clear that maize gel experiments would be feasible, the question of the appropriate glassware became the next challenge. All of the initial experiments were performed in various scavenged glassware, including small 1L cylindrical flasks, large 4L heritage jars and even a 5 gallon glass carboy made for brewing beer. Although these, especially the 1L flasks, were perfect for the preliminary experiments and proved invaluable to developing the various protocols, they all had major flaws.

First, all of the above glassware had many imperfections, mostly occlusions and smears in the glass. Additionally, all glassware used so far was made using a molding process. This involved making several separate parts and then fusing them together. Although this made it cheap to produce, it also resulted in seams where the different parts were fused. This, along with the other imperfections, warped and distorted the root portions of the image being imaged. I needed higher quality glass containers, ideally made using an extrusion process. This would both eliminate the seams, as well as other imperfections in the glass.

Ignoring the quality of the glassware, there was also a second concern. The preliminary experiments showed the maize root system to be quite expansive and fast growing. The roots of maize I planted in the 1L test jars would hit the edge of the jar in a day, while the 4L heritage jars would be outgrown in 2 or 3 days. I found that commercial glassware that was both high quality, as well as large enough to house a maize plant for around two weeks of growth was not available.

Consequently I worked with a glass manufacturer, Quark Glass, to design and manufacture a glass container tailored specifically to growing maize. The jars were manufactured from a long extruded glass pipe that was cut into several sections. A glass bottom was fused to one side. The resulting jars were 45 cm deep, had an outer diameter of 300 mm and a volume of 28L. A single jar cost about $350, most of which went towards

the cost of the glass pipes. Fortunately a single company was already manufacturing pipes of this diameter, or none of the maize experiments would have been possible. A smaller version of the jars was also made for preliminary experiments to alleviate some of the labor required to prepare the large production jars.

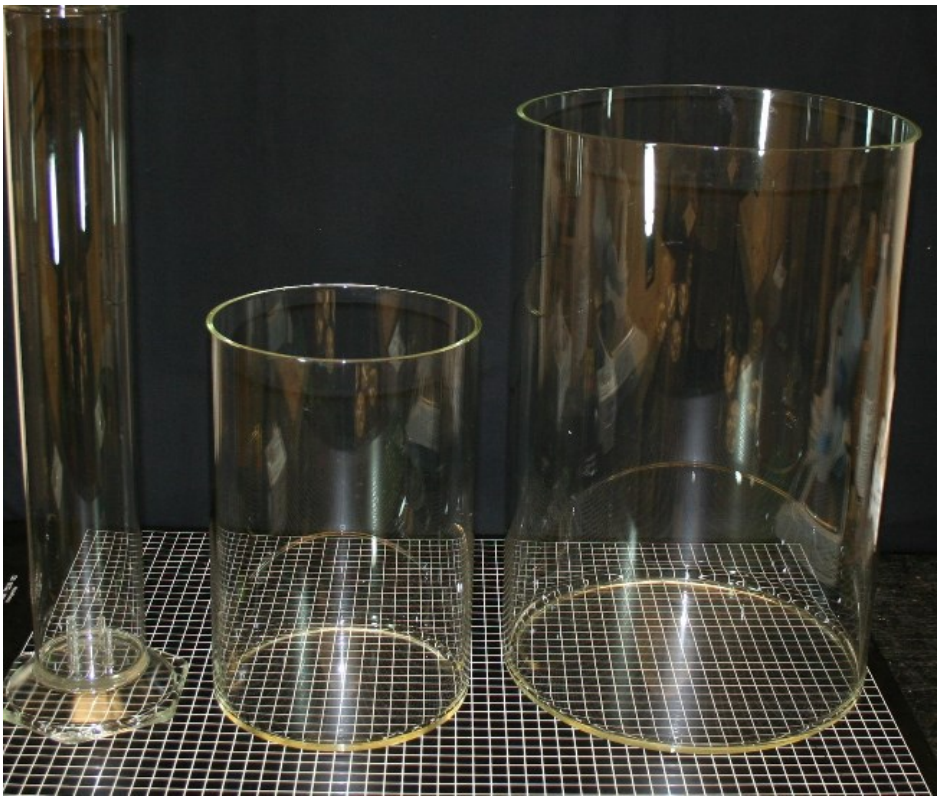


Figure 4: Various types of containers used for gel experiments. From left to right, rice container, test container, full sized maize container.

Growth container and media preparation

As stated previously, the maize jars had a maximum volume of 28L. As the shoot portion of the plant tended to occupy about the same volume as the roots when grown to for 2 weeks, the space in the jar was divided in half. The top half was left empty, while the

bottom half was used to hold 14L of media. The media itself was prepared in the jar using the following steps:

1. An empty jar was placed on a large stir-plate and a stir bar placed inside.

2. 14L of ddH20 were added.

3. 28 mL of Hoagland's solutions A and B were added. (Table 2)

4. 2.8mL of FeDPTA was added.

5. The pH of the media was adjusted to 6.0 using 5M KOH.

6. 21 grams of Gelzan were added and mixed into the solution. This produced a 0.15% gel, which allowed for the greatest clarity while still giving a firm matrix for the roots to grow in.

7. The jar was capped with a double layer of heavy duty aluminum foil, and sealed around the edges using autoclave tape. Two spots, on either side of the jar, were left open to allow steam to escape.

8. The jars were autoclaved in groups or two or three for 90 minutes. Such a long autoclave time was necessary to heat the large volume of liquid to a point where all of the gelzan would melt.

9. The jars were placed at room temperature for about 24 hours to cool and solidify.

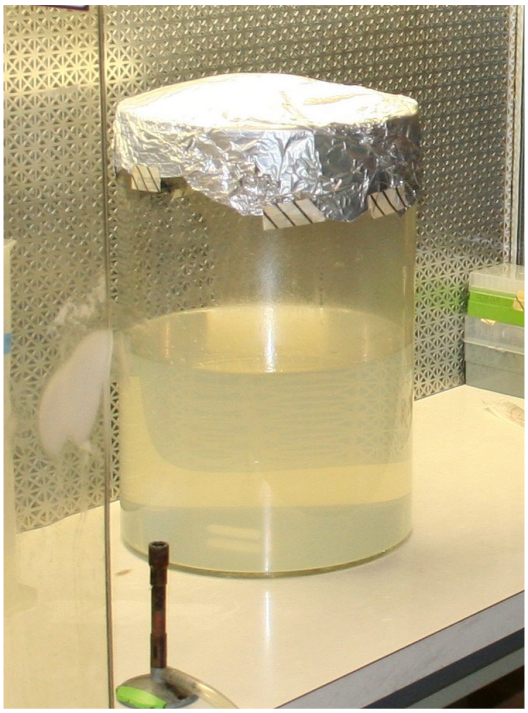


Figure 5: Photo of a full size maize growth container.

Basic imaging setup

In addition to adapting the gel system to maize, the imaging setup itself also had to be redesigned to accommodate the large maize jars and because the imaging setup was too inefficient for large experiments.

The basic imaging setup used for the first gel experiments (Iyer-Pascuzzi et al., 2010) consisted of an Ortery Photocapture 360 turntable (http://www.ortery.com), computer software that came bundled with it, a digital SLR camera (Canon Rebel XT) mounted on a tripod and a laptop. A jar to be imaged was placed on the turntable, while the camera was placed in front of it. The Ortery software would then control both the

camera and the turntable and take a rotational set of images. Lighting was provided from the sides using two swing lamps. These were later replaced with a large light box placed behind the setup (Fig 6).

Figure 6: Basic gel system imaging setup.

This setup would not work for maize as the Ortery turntable was not designed to hold the 40 pound weight of the maize jar. With the help of the Duke Physics Machine Shop I built an aluminum frame that enclosed the Ortery turntable and took some of the weight off the turntable. This was partially successful. There was still too much weight for the turntable to start spinning on its own, but if a bit of pressure were applied to it, it was able to start and keep on spinning on its own power. Although not ideal, it allowed me to collect some preliminary images of maize grown in the new jars and further showed

the feasibility of using the gel system for maize. These preliminary images also demonstrated a problem with imaging cylindrical containers. Filling the jars with media turned the jar into a lens. Images of roots were distorted, especially around the edges of the cylinders, where the curvature of the glass was especially pronounced. The same, but much less pronounced, was true for the jars used for rice experiments.

The solution to this problem, described by (Clark et al., 2011), was to place the jars in a square acrylic tank filled with water. Because the refractive index of the water in the tank, the media, and the glass were very similar, this effectively eliminated the curvature of the glassware. Because the turntable could not be placed inside the water tank, the motion from the turntable had to be indirectly transferred to the jar. This was done by placing a plastic lazy Susan inside the water tank. The turntable was placed under the water tank, which itself was being supported by an aluminum scaffold to prevent the weight of the whole setup resting on it. A set of very strong neodymium magnets were placed on both the lazy Susan and the turntable. As the turntable spun, the magnets placed on it would pull the magnets placed on the lazy Susan, transferring the motion to the lazy Susan, and a jar placed on top of it (Fig 7).

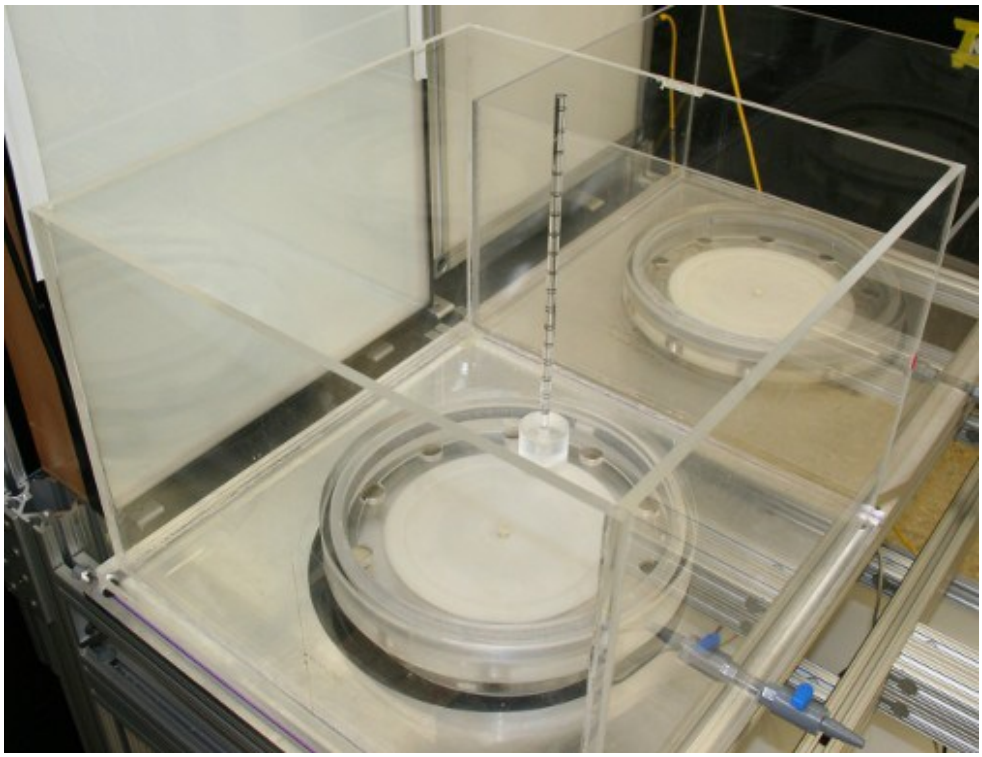


Figure 7: Square water tank and lazy Suzan system.

All of the problems discussed above had to be resolved, and the improvements implemented were significant, but the key issue still existed, and drove the complete redesign of the imaging setup into a much more streamlined design. This issue was calibration. Before each imaging session, the camera had to be calibrated to the rest of the setup. This involved insuring that the camera was completely perpendicular to a sample jar, that the camera was the correct distance from the jar, and most problematic, that the rotating jar would always stay in the middle of the view of the camera and not wobble from side to side. This required a lot of minor alterations to the position of the camera and

could take hours. To make matters worse, the images were taken in a room adjacent to a machine room. As a result, the vibrations from the heavy machinery would transfer to the tripod with the camera, causing it to shake. This caused some of the images taken to be out of focus, and required retaking of whole image sets. The vibration could also throw off the calibration, which then had to be redone. To solve all of these problem, Chris Topp and I developed and built a dedicated imaging platform (Topp et al., 2013).

Improved imaging setup

The dedicated imaging platform was modeled on the old imaging setup, incorporating all of its components and recent improvements (Iyer-Pascuzzi et al., 2010; Clark et al., 2011; Galkovskyi et al., 2012). Additional thought was put into the design to minimize the frequency of calibrations, as well as reduce vibrations by isolating the various components from outside influences.

This was accomplished by turning the imaging setup into a permanent imaging platform. All of the various components required for imaging were mounted to a single heavy aluminum frame. By using a dedicated cameras, only a single thorough calibration was required, and recalibrations were only required when the cameras had to be adjusted for different sets of experiments, or further modifications to the imaging platform had to be done. This has proven to be a very infrequent occurrence, as there were only two

iterations of the platform, the original design, and an upgrade performed a year after the original was assembled.

Several improvements were made during the assembly of the first imaging platform. A minor change was to use a desktop computer instead of a laptop. The performance of the laptop was a minor issue with the simple setup, and using a simple mid-range desktop gave a much needed "quality of life" improvement. Although ultimately all of the images taken were uploaded to a server for processing and analysis, a redundant local backup to an external hard drive was also performed after each imaging session.

The second improvement was to replace the aging Canon Rebel XT camera with a higher end Canon EOS 5D with a 60mm fixed lens. A polarized light filter was also used to reduce the amount of glare. A higher exposure time was used to offset the light lost due to the use of a filter. The new camera has allowed us to take higher resolution images, which improved the clarity of images, which in turn made all subsequent image processing much easier. This was especially helpful for maize experiments, as the field of view of maize images was larger due to its larger root system.

The camera itself was mounted on a movable aluminum tower attached to the aluminum table frame (Fig 8). The tower could be moved towards or away from the tank, as well as side to side. The camera was attached to a mount on the tower, and the mount

itself could be moved up and down. This freedom of movement in all 3 axes allowed us to position the camera in an ideal location so that we could capture the whole root system at as high resolution as possible. In practice that meant that the camera tower was pushed as far back as possible, and then a fixed focal length lens was selected such that most of the field of view was taken up by the jar. The forward/backward position of the camera was then further adjusted so that the left and right edge of the jar lined up with the left and right edge of the field of view of the camera.

A 60mm focal length lens was used as it allowed us to move the camera almost all the way back to the end of the table. A variable focal length lens could have allowed us to move the camera even further, but the increase in resolution would have been minimal, and variable focal length lenses tend to trade image quality for versatility. As the imaging distance would be constant once set, a fixed focal lens ideal for our particular distance was a much better choice.

The Ortery turntable was set on a shelf at the far end of the platform. The aluminum support was placed over the top of it, and the water tank was placed on top. As previously described, this aluminum support held the weight of the water tank and the jar being imaged, while the turntable, via the magnets, would rotate the platter (lazy Susan) inside of it. Lastly a fluorescent light box was hung behind the water tank to provide illumination (Fig 8).

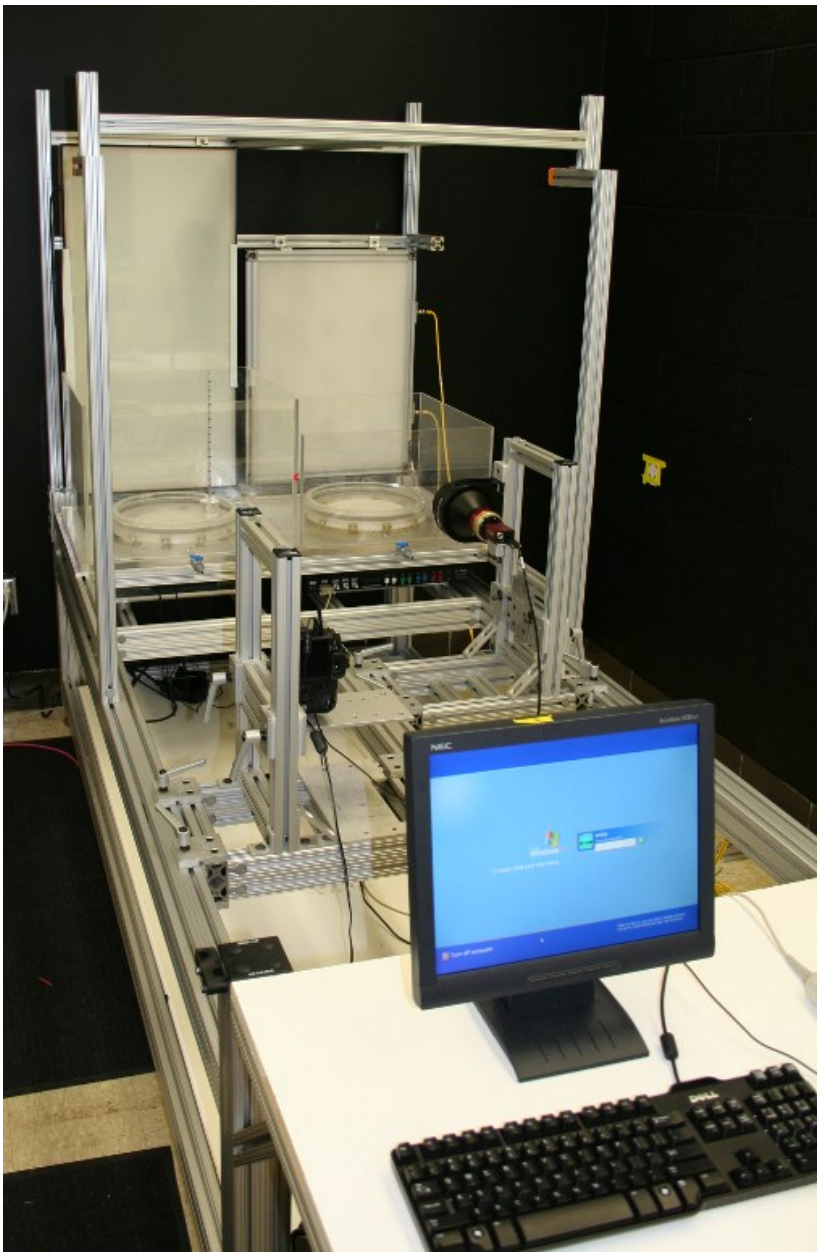


Figure 8: Improved imaging platform. Left setup consists of a fluorescent light box, a Canon camera and an Ortery turntable. Right setup consists of an LED backlight a Stingray camera and a LinerX heavy duty turntable.

Further improvements of the imaging setup

The imaging platform described above was used for the majority of the imaging done in Chapter 3. After a year of use, the Ortery turntable became less and less reliable, and finally broke. This was likely because most of the components of this turntable were plastic and not designed for the heavy duty use we put it to. Instead of replacing it with the same model, we decided to upgrade to a more robust industrial turntable. This turntable, LinearX LT360EX, uses metal components including a steel worm drive (http://www.linearx.com). As a result not only was the turntable rated to handle significant weight loads, a preliminary test of the table allowed me to take a rotational series of images of myself, but also allowed for much greater precision of movement. Additionally it allowed for much finer controls of both speed and acceleration. Unfortunately, the Ortery software was designed to only work with the Ortery turntable. Although the LinearX turntable did not come with any usable software of its own, it did come with an extensively documented API. This allowed us to create imaging software, through a LabView software contractor, specifically tailored to our needs.

Additionally, the Canon camera was replaced with an OEM digital camera, an Allied Vision Stingray F-146B/C with 35mm lens (http://www.alliedvisiontec.com). This camera, although monochromatic, uses a CCD sensor which produces slightly higher quality images. This camera lacked a lot of the bells and whistles of a consumer camera

and was simple small rectangular box with the sensor mounted inside. Because of that it was much easier to mount to the camera tower. Although the camera change increased the quality of the images, the main reason for the change was driven by other factors. The need to creating our own imaging software for the new turntable would require implementing software for the Canon camera. Unfortunately the Canon camera we were using was very hard to control using 3rd party software. The Stingray, on the other hand, had a very simple and very well documented API which ended up working extremely well with LabView.

The last major component replaced during this upgrade was the light source. The light box used so far consisted of a wooden box with two long fluorescent bulbs. The light was passed through a diffuser on its way out the box to distribute it more evenly and eliminate the appearance of the two bright lines where the two bulbs were located. This worked to a certain degree. While the light was mostly uniform in the middle of the box, there was still a light gradient towards the edges of it, with the light at the very end being noticeably dimmer.

This had an interesting effect when combined with the varying thickness of the gel (thin at the edges of the jar, and thick through the middle). The part of the image of the bottom middle of the jar, where the light was dimmest and the gel thickest, was the darkest. Going up the image, as the light increased, increased the lightness of the image.

Going outward in the image, as the thickness of the gel increased, also increased the lightness of the image. The result was a conical gradient of contrast in the image, which consistently introduced a thresholding artifact during image processing. Image processing will be fully described in detail in the next section, but one of the first steps in image processing is thresholding. This is necessary for automated phenotyping, as the algorithms involved need implicit knowledge of which part of the image is root and which is background. Thresholding is the process where a color or grayscale image is converted into a simple black and white binary image. The value of all background pixels is set to black, while the value of the root pixels is set to white. At its simplest, this is done by setting a color threshold. Everything darker than some intensity value is consider root, while everything lighter is considered background.

A high contrast between the background and the root system would allow for the easiest thresholding. Consequently, lack of contrast would produce artifacts, and roots close in color to background would either be treated as background itself and ignored, or worse, the background itself would be considered part of the root. The gradient described above resulted in the latter. The well-lit portions of the root systems were easily separated from the background, but because of the lack of contrast in the gradient, the whole of it was also treated as root. The resulting thresholded images consisted of the root system at the top, and attached to it near the bottom, a cone shaped artifact. Some of this was

resolved by optimizing thresholding settings, and will be discussed further in a later section of this chapter.

As the replacement of the turntable and camera would require heavy modification to the table, we also took this opportunity to implement a hardware solution for the light/thresholding problem. We replaced the fluorescent light box with a LED panel. The light in these types of panels is provided by a large array of individual LEDs. Combined with a diffuser, the resulting sheet of light was much more uniform in intensity. Additionally the LED panel was much brighter than the fluorescent light. The combination of these two factors almost completely eliminated the gradient of light. The remaining non uniformity in background intensity were easily dealt with by optimizing thresholding parameters.

Lastly, because most the major components of the table were replaced, minor changes to the frame itself also had to be made. These were mostly done because the new components were of different size or shape as compared to the old ones. An aluminum mount was made to allow mounting of the camera to the camera tower. The shelf which held the turntable was replaced with a sturdier one, and a permanent water tank support attached to the whole frame was added right above the turntable. The lazy Susan plate inside the water tank was replaced with a larger one with more mounts for magnets. Additionally the plate was made with a recess so that the jars would fit snugly into it. This

ensured that when a jar was placed onto the plate to be imaged, it would always be in the center of the lazy Susan. Lastly, the light box mount was adjusted to hold the new LED panel.

Camera calibration

As previously mentioned, camera calibration was a necessary, but extremely time consuming part of any imaging session. Calibration was done by eye using trial and error. Although 2D analysis did not seem to suffer due to bad calibration, 3D reconstructions were much more sensitive to it. The new imaging platform solved some of these issues. Calibration only had to be done once since all the components were locked in placed and any further recalibrations only had to be done if the setup had to be readjusted. Most importantly, we also replaced the trial and error method with a standardized protocol.

An ideal image set is one where the left and right edge of the jar line up with the left and right edge of each image. In order for this to happen the jar being imaged had to be in the middle of the lazy Susan plate while the camera had to be pointed at the horizontal midline of the jar. More correctly, the camera had to be pointed at the axis of rotation of the turntable, but as long as the jar was placed in the middle of the plate, the center axis of the jar would line up with the axis of rotation of the turntable.

In order to calibrate the setup, the various components of the platform were moved around until an ideal image set could be produced. In most cases this could be

accomplished by moving the camera alone. Bad miscalibration resulted in a "wobble" of the jar. As each successive image was taken, the image would drift to the side, and then return back to the center as the jar came back to its original position. At its apex, this wobble could place part of the root system completely out of the field of view. By ensuring that the jar was placed in the middle, and that the camera was placed correctly, the wobble could be eliminated. To ensure that the jar was placed correctly we used the recessed plate described above. Because we had to leave some room for the jar to be placed into the plate, a very small amount of wiggle room was available. Fortunately this only introduced a tiny amount of wobble and even this could be prevented by carefully placing the jar in the middle of the recess.

The second way to introduce wobble was improper placement of the camera. Ideally, the very middle of the field of view of the camera had to line up with the axis of rotation of the turntable. If the camera was too far left or right, or not pointing straight at that line, the rotation of even a perfectly centered jar would result in wobble. Camera calibration should have been possible by ensuring a sample jar was exactly centered, and then taking successive rounds of images, while adjusting the position of the camera until the wobble was gone. This in fact was the approach taken with the original imaging setup. Unfortunately it depended on a sample jar being a proxy for the axis of rotation of the turntable. The small amount of wiggle in the recessed plate made it less desirable.

A much better proxy was designed and made for us by Tim Horn (Fig 9). This proxy consisted of a perfectly machined round aluminum cylinder. Protruding from the middle of the cylinder is a tapered aluminum spike. The goal was to place this proxy such that the central axis of the spike lined up with the axis of rotation of the turntable. To accomplish this we used a dial indicator (Fig 9), which allows for accurate measurement of distances.

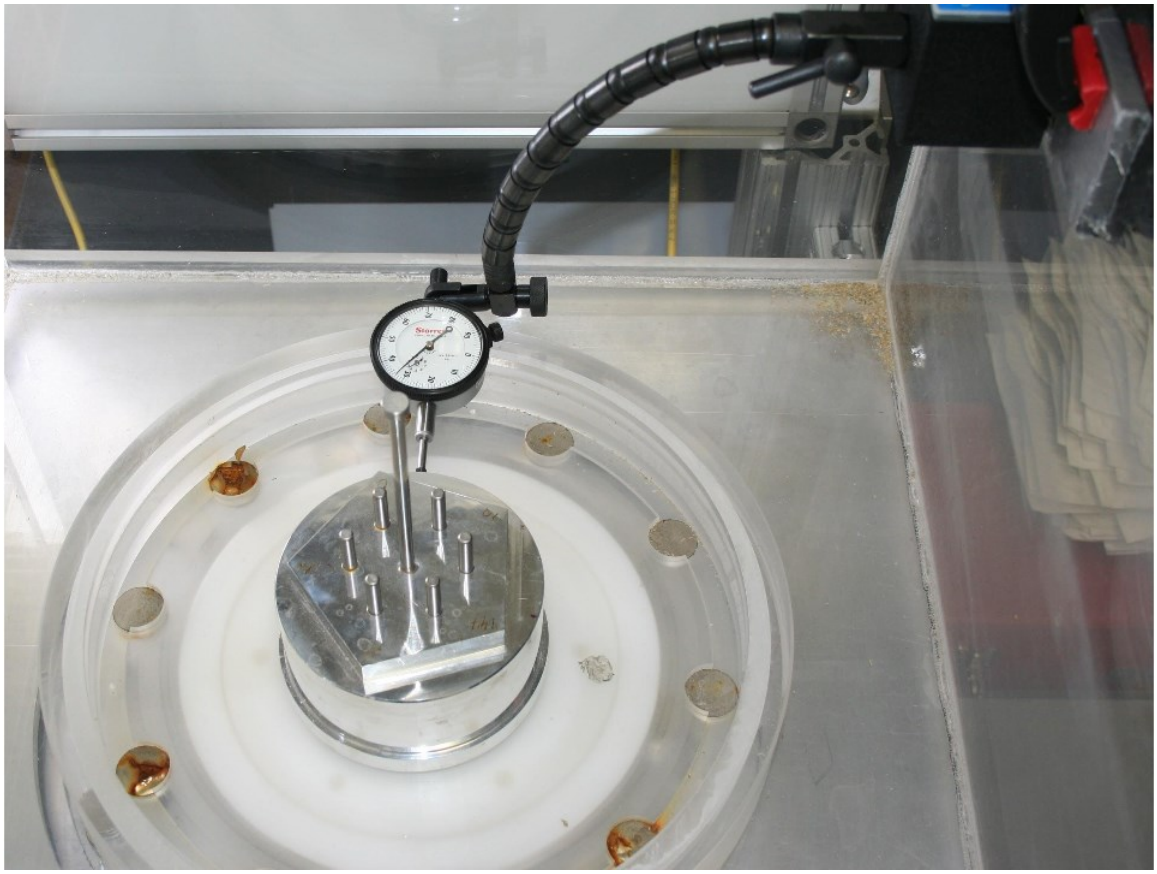


Figure 9: Calibration of the imaging platform

This is best described by example. Let us assume that the proxy object was placed incorrectly and it was slightly off center. Let us further assume that you can measure the distance from some fixed point in space, to a point on the proxy. This is not a fixed point, but for an example, a point that corresponds to one of the cardinal directions. As the proxy spins, if you keep track of the distance between your fixed point, and the point on the proxy that corresponds to say North, that distance will change if the object wobbles. On the other hand, if the proxy is exactly in the middle, the distance will never change. Adjusting the proxy so that the distance remains the same, lines up the axis of rotation of the proxy with the axis of rotation of the turntable. The dial indicator is a little gadget that allows measurement of that distance. By temporarily attaching it to the water tank we establish the fixed point while the dial shows any variations in distance. The proxy is then gently moved around until centered. The camera is then adjusted such that the aluminum spike protruding from the proxy is always in the center of the field of view. The imaging software has an option to show gridlines for this purpose.

Image processing and analysis

Following an imaging session all of the images are uploaded to a local server as well as backed up to an external hard drive. The image names follow a standardized format. An example would be something like this: ZmNAMp0001d04_001.png. The first two letters indicate the species, Zm for maize. The next 3 letters make up the experiment

code. Next comes the letter p followed by a plant number, which is followed by the letter d and the imaging day. Lastly comes an underscore and the number of the image in the series. The species, experiment and plant number make up the plant ID number which, when cross referred to a master spreadsheet, indicates the germplasm in question.

Once uploaded to the server, a script parses the name of each file and sorts all the files into appropriate folders. These folders are arranged hierarchically, with a corn and rice folder on top. These contain the various experiment folders, which in turn contain individual plant folders. Lastly each plant folder contains individual day folders with the corresponding set of images. This makes finding any particular plant extremely fast and easy.

Image processing and analysis is done using several different programs. These applications have been coded in house or by collaborators. The full list, including the creators and available references are listed in Appendix A. These application have to be run in a fairly linear order, as outputs of most serve as inputs for only a single other application. Because of this, we've created a single pipeline GUI that allows access to all of the applications from a single place as well as guides the user through the analysis process.

The first step in the pipeline is setting the scale for the images. This is the ratio of pixels to millimeters and allows for conversion between the two. The scale can be set in

two different ways. If the scale is already known it can be set directly. If the scale isn't already know, a reference of known size has to be used. In our case the size of a jar itself was used, as its outer diameter was known to be 300 mm. The scaling application asks the user to draw a line on the reference, in our case this was a line from one edge of the jar to the other, and indicate its length. During data export, this ratio is used to convert all pixel values into millimeters. Once obtained, the scale remained the same until the camera was replaced. As such, the scale for all of the images taken with the Canon camera was .0965 mm/pixel, while the scale for the Stingray camera was .1235 mm/pixel. All of the data presented here has been appropriately scaled.

The second step in the pipeline is cropping the images. Sections of the raw images contain the shoot portion of the plant, as well as the edges of the cylinder, are removed. The sharp changes in color at the air and gel as well as the gel and jar edge interface are currently unable to be detected by our software and would interfere with further image processing. As such, the images have to be cropped so that only the root system remains. This is a two-step process. First the user is shown a single image from a set and asked to draw a top crop line. I set this slightly below the level of the gel. This same top crop line will be used for all of the images in a set. The second step is to draw the crop box around an image, as would be done in any type of image editing software. At this stage the top crop line is fixed in place, but can be unlocked if necessary.

All of the images have to be cropped to the same size, and so the rectangle has to be selected such that, no matter which image in the series is cropped, all of the roots will be included. Because of this a single image cannot be the basis of selection for the crop box. For example, an outlying root that points straight right in a first image of a set, would be pointing straight left in the 21st image. A crop box selected based only on the first image when applied to the 21st would crop away the outlying root.

In order to see the furthest possible extent of the root, the cropping is done using a unique composite image of all the images in a series. This image is created by looking at the color of any given pixel in the same position in all images in a set. Because roots are always much darker than the background, the darkest value is selected and used in the composite image. The resulting image shows all of the roots in a single image. This allows us to draw a crop box that, when applied to the individual images, will be certain to include the whole root system.

The third step in the pipeline is to perform 2D analysis of the images using the GiaRoot application (Galkovskyi et al., 2012). This application transforms the images into a software algorithm usable form, i.e., thresholds them, and then performs 2D phenotyping. This analysis is done semi-automatically, as the only user input involves selecting thresholding parameters. Depending on the type of thresholding used, several parameters have to be fine-tuned to produce the best thresholding. All of the images used

here were thresholded using “double adaptive” thresholding. This algorithm works by growing a square around a pixel and looking at the change in intensity in the area covered by the square. Because the roots are much darker than the background, a change in intensity will only occur in a presence of a root. One of the parameters, neighborhood size, controls how large the square is allowed to grow, while a second, bound drop value, controls how strong the change of intensity has to be in order for the pixel to be considered a root pixel. The last parameter, max component size to ignore, determines maximum size of components to be ignored. This parameter eliminates any small components. These are the results of small bits of floating detritus that accumulate in the tank over an imaging session and are generally not connected to the main root component. After thresholding, the GiaRoot software calculates values for a set of 17 2D traits (Galkovskyi et al., 2012; Iyer-Pascuzzi et al., 2010).

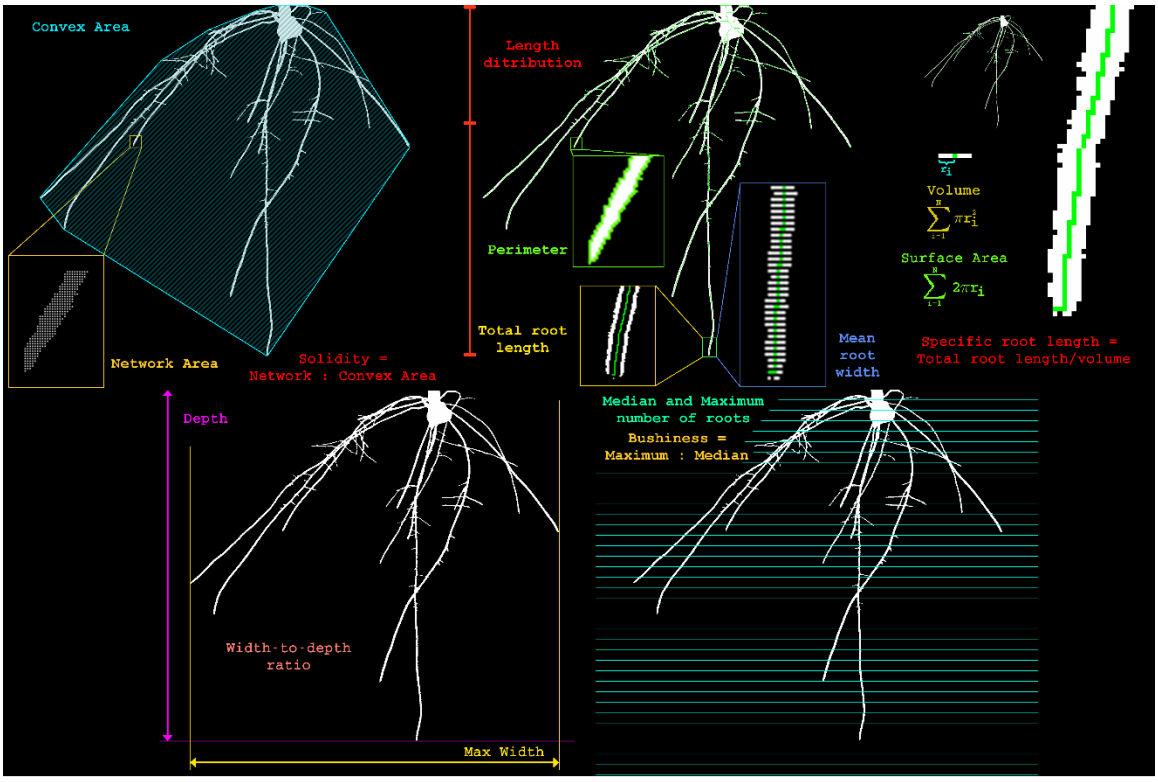


Figure 10: 2D traits calculated by GiaRoots.

Selection of good parameters is a key step in image processing. If the thresholding is not stringent enough, noisy images will be produced. The roots could have jagged edges as background pixels can be counted as root pixels, or the small artifacts produced by the floating debris could be retained. On the other hand, if the thresholding is too stringent the roots can end up thinner than they should be, as some of the lighter root pixels would be treated as background. Parts of, or even whole roots could be eliminated altogether. The finer roots that do not contrast with the background as much tend to be especially susceptible to this.

At this stage, the 2D trait data can be exported and used in subsequent analysis, or an optional step can first be performed. This step involves performing quality control (QC) of the thresholding. Because thresholding is such a critical step of the pipeline this step is highly recommended, and all of the data presented here have gone through it. The quality control involves manually checking all of the images for visual artifacts or any other inconsistencies. On average about a third of any given batch of image sets failed the first pass through the QC, but a majority of these were easily fixed.

The QC is performed using a Python script accessible through the pipeline GUI. This script takes all of the thresholded images of a given set, scales them down, and then arrays them in a 4 by 10 grid. This allows the user to see all of the images in a set in a single glance. Furthermore the script allows the user to tag the set as either having passed or failed the QC. If failed, several additional tags can be applied that describe the nature of the failure. These are used to guide the approach to correcting the thresholding.

The most common problem was caused by incorrect cropping. The top crop line was placed too close to the surface of the gel or the side crop. This is easily fixed by recropping, either using the original images, or using the recrop application which allows to crop the already cropped images.

The second most common problem was appearance of artifacts. This was resolved by using a different set of thresholding parameters with a larger value for the max

component size to ignore. Less frequently, and usually for young roots, some of the images had no root system at all. This was due to the root being so small, it was completely eliminated due to too large of a value for max component size. This was easily fixed by using thresholding parameters with a smaller max component size to ignore. Once all of the thresholding issues had been resolved, the 2D trait data were exported into a spreadsheet and further analyzed. These analyses are further described in the two subsequent chapters.

It is worth mentioning that the analysis pipeline has several additional capabilities. These are reconstruction of a 3D model from the image sets followed by phenotyping using a set of 3D specific traits. The prohibitive factor for these analyses was the reconstruction itself. Although quite robust for rice roots, the software was unable to reconstructs about half the samples for the NAM founder experiments. This is most likely due to the larger size of the corn root. In retrospect image sets composed of larger number of images could have helped, but fortunately the 3D traits were, for the most part, not required as demonstrated in the subsequent section on system validations.

Ground truth system validation

In order to validate that both the mechanical and software aspect of our gel system were functioning as expected, we performed several validation experiments. We wanted

to ensure that the trait measurements calculated by the pipeline were realistic and reflected the true measurements of the root system being imaged.

Initially this proved to be somewhat tricky. Ideally we would have wanted to obtain true measurements of a root system, without using the gel system, and then compare those to measurements produced by our phenotyping software. Unfortunately we did not have an a priori method to determine the true measurements of a root. Some simple measurements, such as depth or width, could theoretically be taken, but most of our traits would be quite difficult to measure by hand.

Fortunately a validation model could be used instead. This could be anything so long as it could be easily measured by hand and subsequently placed and imaged in the gel. Several such objects were used. Initial validation used a facsimile root made out of wires of various thicknesses (Iyer-Pascuzzi et al., 2010). It demonstrated that the phenotyping software was quite robust, but because the model was complex and required hand measurements, only four of the simplest traits: depth, max width, root length and average root radius, could be validated.

I wanted to use something that would validate all the traits. My first attempt involved using much simpler models, a musical tuning fork and a magic linking ring. Although these were much easier to measure by hand, they were still unsuitable for validation of complex traits, but proved very useful in validating 3D reconstruction

software during its development. Because of the complexity of the root system, errors in their reconstruction could be overlooked, but because the ring and the fork were such simple constructs, errors in their reconstructions were easily noticeable.

At this stage, it became apparent that in order to fully validate the system, I had to have access to a validation model whose measurements I knew, preferably in the form of our 2D and 3D traits, without actually measuring the model in the first place. The trick to doing this was to start with a set of measurements, or more specifically a digital version of a model, and then manufacture a model to fit those.

To begin with, I wrote a Python script that created a 3D voxel model of a root-like object (Fig 11). This was done by building several cylinders of voxels by stacking circular layers of voxels on top of each other. If the layers were stacked on top of each other, the cylinders would be completely vertical. If the layers were offset from each other, the resulting cylinder would be angled. Varying the direction in which the layers were offset resulted in cylinders going in different directions, while regularly interspacing offset layers with ones that weren't offset resulted in the cylinders having different angles. Varying the diameter of cylinders produced projections of several different thicknesses. Lastly, I virtually rotated this model while taking a set of 2D images, and ran both through our analysis pipeline to produce a set of 2D and 3D traits. Importantly, no hand measurements were taken and the values for all traits could be generated.

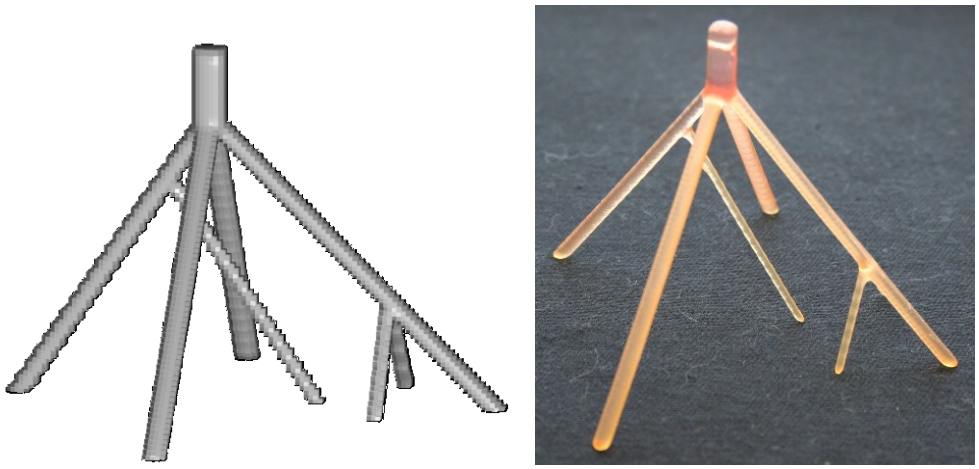


Figure 11: Figure of the digital (left) and 3D printed (right) validation model

The next step was to manufacture a physical version of this model. This was done by FineLine, a rapid prototyping company. They were able to create a physical model from the digital one I provided (Fig 11). This was done using 3D printing, which involves successively placing very thin (100 um, below the resolving power of our cameras) layers of plastic on top of each other. Because the layers are so thin, the resulting physical model was an extremely good representation of its digital counterpart. This model was then imaged, phenotyped, and the resulting traits compared to the ones produced from phenotyping the digital model (Table 2) (Topp et al., 2013).

2D traits	
	%difference
bushiness	2.63
perimeter(mm)	0.30
majorellipseaxes(mm)	-0.27
networkarea(mm^2)	1.68
minorellipseaxes(mm)	-0.17
widthdepthratio	0.55
depth(mm)	-0.45
maxwidth(mm)	0.10
networkconvexarea(mm^2)	0.48
ellipseaxesaspectratio	0.09
solidity	1.24
maximumnumberofroots	-1.54
mediannumberofroots	-4.35
averagerootwidth(mm)	-5.99

Estimated from 2D	
Trait	%difference
Totallength (mm)	27.88
Specificrootlength (1/mm^2)	7.20
Volume (mm^3)	68.33
Surfacearea (mm^2)	19.93

Calculated from 3D model	
Trait	%difference
TotalRootLength 3D	-0.91
SpecificRootLength 3D	10.67
RootSystemVolum 3D	15.28
SurfaceArea 3D	14.24

Table 2: Table of the accuracy values calculated by comparing the trait values derived from the digital and 3D printed model.

These data showed that a large portion of our traits are measured correctly and are only a few percent off. A few traits are measured incorrectly using 2D phenotyping but are more in line with the true values when using 3D phenotyping. Interestingly those traits that are far better in 3D, volume or surface area, are actual 3D measurements. This is likely because our 2D volume and surface area are modeled estimates and not direct measurements. It would make sense that using a 3D approach to measure a 3D trait would produce a much more reliable result.

Chapter 3. Variation of RSA among 26 diverse maize lines

Once the imaging system was up and running, I begun tackling biological questions. As described in Chapter 1, root studies of crop plants are a small, but increasing, portion of plant research. One of the goals of developing the gel imaging system was to use it to locate loci controlling root system architecture (RSA) through quantitative trait loci (QTL) mapping experiments. At its most basic, QTL mapping involves genotyping and phenotyping a mapping population followed by statistical tests which determine if any loci in the genome are linked with the phenotypes measured (Lander and Botstein, 1989; Dupuis and Siegmund, 1999; Reiter et al., 1992). Although several different types of mapping populations can be used, the Recombinant Inbred Line (RIL) populations are the most commonly used (Bailey, 1971; Crow, 2007; Yu et al., 2008). These are created by crossing two different inbred lines with contrasting phenotype or highly distinct genetic backgrounds. A single offspring is then selected and selfed (Fig 2). Around 150 to 200 F2 progeny are then propagated through single seed descent for 6 or more generations to produce the final RIL population.

This accomplishes two goals. Crossing of the two inbred founders allows their genomes to mix through subsequent rounds of recombination, while the repeated rounds of selfing causes the fixation of these new recombinant genomes. Although 6 rounds of

single seed descent fixes most of the genome of each line, some residual heterozygosity remains. In the NAM population this ranged from 3.2% to 4.1% (McMullen et al., 2009). The resulting population is composed of a number of individuals, the genome of which is a different mix of the two parental genomes. The different genomes result in phenotypically different plants, which can be analyzed by QTL mapping programs such as QTL Cartographer, to locate the underlying loci (Basten et al., 2004).

The inbreeding required to create the RIL population has the added advantage of immortalizing the population. This makes it a great resource to the community at large as, once created, the population can be used over and over by different researchers, such as in the case of the popular B73 x Mo17 (IBM) population (Lee et al., 2002). Because this would eliminate a lot of the time and resources required to create our own, we've decided to select one of these for our work. Unfortunately little data were available that described the root system of founders of the different RIL populations. In order to make an informed selection, we decided to perform a survey of the inbred lines which were used to create the NAM RIL population. This would also serve as a survey of root system architecture of maize.

Selection of maize varieties for screening

Although several different maize RIL populations are available, the Nested Association Mapping (NAM) population was especially promising. This population is a

group of 25 different RIL populations, each composed of 200 individual lines, all of which share B73 as one common parent (Yu et al., 2008; McMullen et al., 2009). The B73 variety was chosen due to its history as an agronomically important line as well as its choice for the maize whole genome sequencing project (Schnable et al., 2009). The remaining 25 varieties were chosen based on 94 microsatellite markers to include as much maize genetic diversity in the population as possible (Yu et al., 2008). The founders include varieties from several different heterotic groups (Shull, 1908; Yu et al., 2008), including the temperate stiff and non-stiff stalk groups, as well as well as the tropical group. Additionally two sweet corn as well as a popcorn lines were included.

Screening the founders of the NAM population offered several advantages. First because the founders were selected to encompass a large portion of the variation in maize, the chances were very good to find phenotypically distinct root systems. Additionally, due to the single common parent, screening the 26 NAM founders would let me choose from 25 different RIL populations, while screening 25 individual lines would require upwards of 50.

Secondly, all of the individuals in the entire NAM population were already genotyped, first using SNP chips (Flint-Garcia et al., 2005) and then by genotyping-by-sequencing (GBS) (http://www.panzea.org/lit/data_sets.html). This meant that after screening the parents we could directly transition to phenotyping the selected RIL

population. Additionally the use of GBS for genotyping produced a much higher marker density than would be available in most other RIL population. The linkage map used here included a set of phased an imputed markers one centiMorgan apart.

Experimental Design

In addition to the protocols outlined in Chapter 2, several additional points had to be considered. First, because of lack of information as to the extent and speed of growth of maize roots we had to determine how often to image as well as well as how long we could grow plants before they ran out of room in our containers. Initial experiments with several different varieties of maize performed during the development of the imaging platform showed that there was enough room for around two weeks of growth for the varieties with the largest root systems. Additionally it took about 2 or 3 days of growth for significant amount of growth to occur. Consequently, for screening the NAM founders we imaged the plants on days 3, 6, 9 and 12 after planting. Lastly we planned on acquiring at least 6 replicates for each variety.

As previously stated, most of the images were taken on the first iteration of the imaging table using a Canon camera. Unavoidably, due to the need to replace the broken turntable, the last few sets of images were taken using the Stingray camera. Luckily, because all of the measurements were scaled and converted to SI units, the use of different cameras did not cause any issues.

Variation of RSA across the NAM founder lines

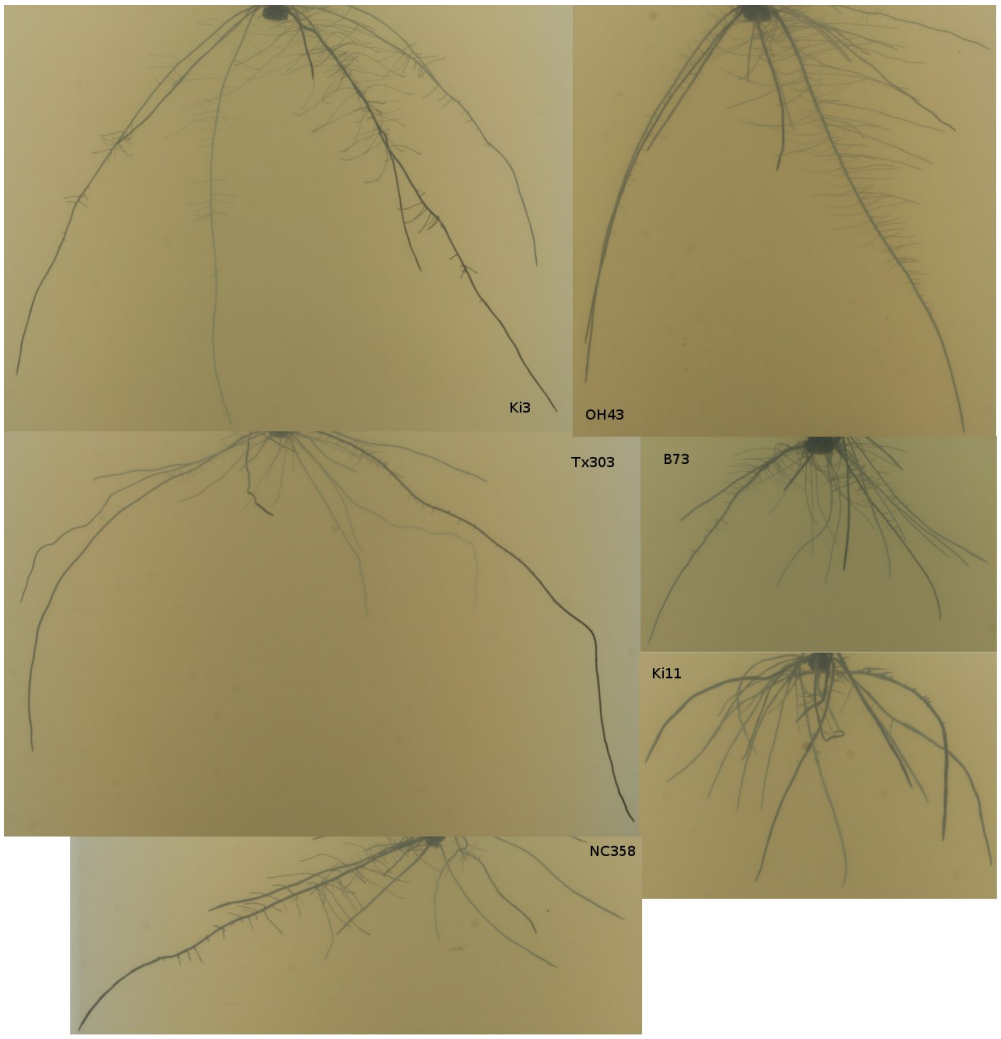


Figure 12: Range of RSA displayed by some of the NAM founder lines. A representative sample for all founders is located in Appendix B.

As expected, there was a wide range of different root system architectures among the different founders. Overall architectures ranged from large and expansive to small and compact with intermediate ones as well (Appendix B). On the large and expansive

extreme were the founders Ki3 and Tx303 (Fig 12). These had the largest values for depth, maximum width, convex hull, major and minor ellipse axes among all of the founder lines. In fact, the Tx303 founder was so large I had trouble obtaining samples on day since in majority of the cases it had already reached the side and/or bottom of the jar.

Conversely, these founders have some of the smallest values for solidity and bushiness. This shows that although these roots encompass a large amount of space, they only explore a relatively small portion of it. This can be seen by eye as large empty pockets of open space surrounded by the outlying roots. This could potentially be a mechanism for scouting out the area to find the best possible location to invest resources, or a way to carve out a territory early with the intent of filling out the enclosing pockets at a later time.

On the other hand, varieties such as B73 or Ki11 had much smaller root systems (Fig 12). The depth, width and convex hull areas were some of the smallest among all of the founders. Because all of the roots of these varieties were packed into a smaller area, the solidity and bushiness trait values were also much larger. This describes plants that forgo initial far reaching exploration in favor of establishing themselves at the site of germination. This could be an advantageous strategy if the seedling is located in a nutrient-rich environment such as in an agricultural setting (Jobbágy and Jackson, 2001).

These two groups, the large and bare versus the compact and dense, make up the two extremes of RSAs seen. The remaining varieties fell in between these two. For

example OH43 had a deep, on par with Ki3 and Tx303, but narrow root system, while its solidity and convex hull area had intermediate values. On the other hand the variety NC358 had a shallow, more so than B73, but quite wide root system (Fig 12).

Interestingly there were no varieties that had RSA that both encompassed a large area while also filling in all of the available space. This seems to indicate limits, at least this early stage, to what the different varieties can do. This could potentially be due to limited amount of biomass, and results in different varieties employing different strategies for its allocation. In cases such as B73, it's to fill in a small area and at the cost of exploration, while in cases such as Ki3 it's to sacrifice root density in order to form a network of wide ranging exploratory roots.

It would be interesting to see how much of an effect these early RSAs have on the RSA of fully grown plants. Most likely the compact varieties will ultimately have to grow more extensive root networks, while the exploratory varieties will at some point have to fill in the empty pockets of space. A deep root system could be especially useful in water poor areas as it would allow the plant to reach deep water (Lynch, 2013), while a shallow root could be especially useful in nutrient rich soils as nutrients, especially phosphates and in fertilized fields, tend to be found near the top of the soil (Jobbágy and Jackson, 2001).

Trait importance and ranking using machine learning

In order to determine the importance of the traits for differentiating between different maize varieties, we performed a machine learning analysis using the mlpy Python library (Albanese et al., 2012). Machine learning works by constructing a mathematical model using data for which you have some kind of metric, in our case, trait values, as well as the class of each data point, in our case the genotype of each individual plant. If enough data is provided in the training set, a highly predictive model can be constructed that can determine the class of any unknown sample (Fan et al., 2008). That in itself was not useful in this case since we already knew the identity of each plant, but the model itself also contained information about the relative importance of the different traits in predicting the class of the unknown sample.

We decided to use logistic regression because this method constructs a simple linear model with a coefficient term for each trait (Fan et al., 2008). This term represents the contribution of that trait to the classification of an unknown sample without requiring any further analysis. Each B73 x "other founder" was analyzed separately and the resulting data are shown in Fig x. High coefficient values indicate that a high value at that trait for an unknown sample would make it more likely to be identified as B73, while large negative values make it more likely to be identified as the other founder. Additionally

traits with the highest positive or lowest negative values are the traits that are key to separating the two founders in that particular model.

The logistic regression was performed with each B73 x "other" founder pair and each day analyzed separately. The data for the pair being tested was randomly partitioned into two even subsets. Half of the data was labeled the training set, while the other was labeled a validation set. A model was constructed using the training set, and used to determine the classes of the individuals in the validation set to test its accuracy. A cross-validation was performed by permuting the above process 1000 times, with the data partitioning being done randomly each time. The resulting accuracies and trait coefficients were averaged. For most pair-wise comparisons, a highly predictive model could be constructed (Fig 13). To further show that the results weren't due to random chance, a control was done by randomly permuting the data labels and running the same analysis (Fig 13). Models constructed in this way could, on average, could not reliably predict. Since there were only two possible classes in each model, this shows that these models had no predictive power at all.

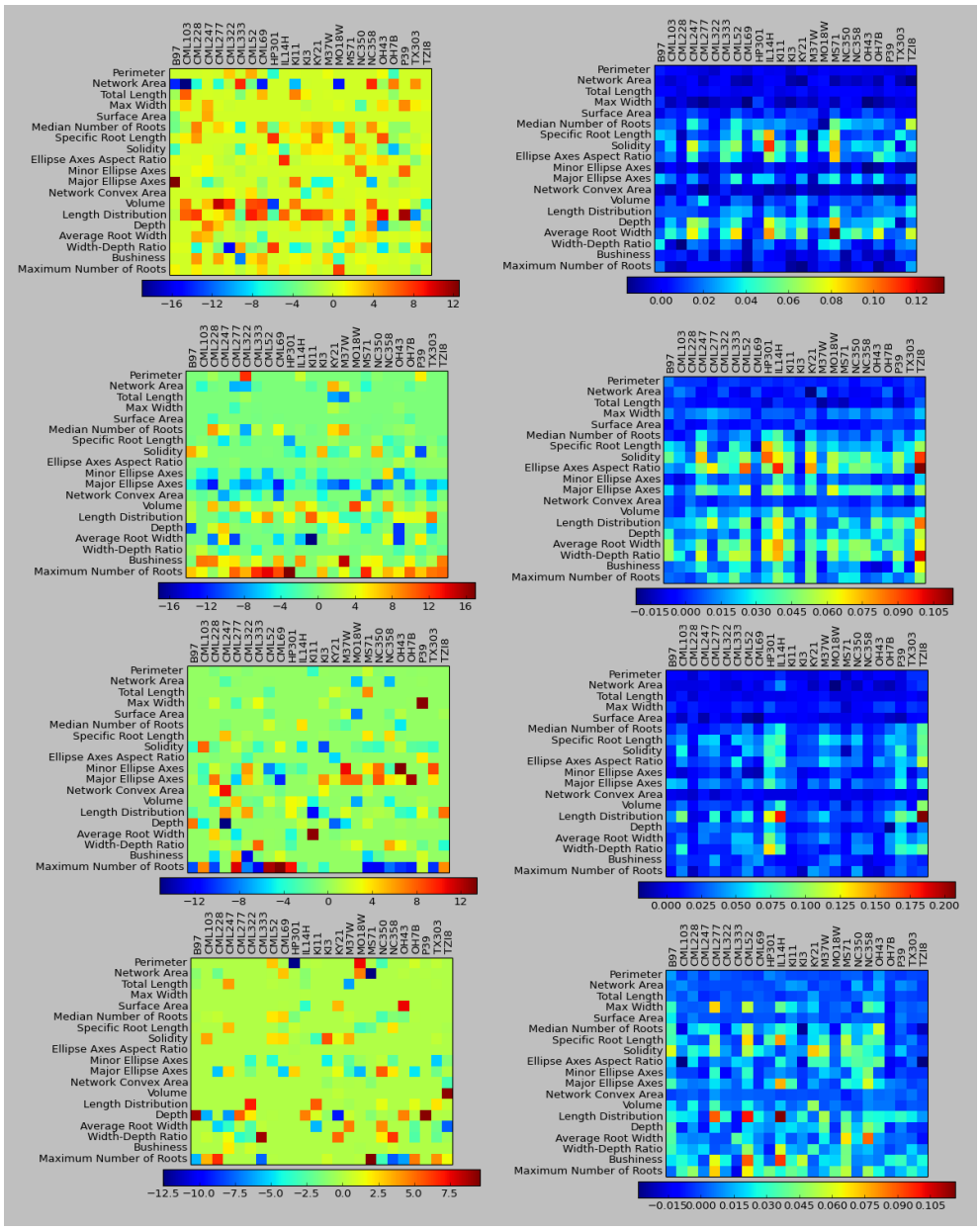


Figure 13: Machine learning figures. Plots in left column represent experimental data. Plots in right column are controls based on randomized data. From top to bottom the rows are day 3, 6, 9 and 12 data.

The machine learning analysis showed several interesting things. First of all, the number of traits needed to accurately classify two varieties was highly variable. For example, differentiating between B73 and B97 could be done using only a few traits. This did not mean that only those traits could differentiate between those varieties, only that those particular traits were the distinguishing ones. Temporarily eliminating the top trait tended to result in the second best trait becoming the top one. Iterative removal of the top trait would repeatedly shift the position of the key trait to the next trait down the line. Traits not used in the full model would begin to be included. Ultimately the predictive power of the model would suffer, but in most cases a large portion of the traits had to be removed.

In contrast to the B73 by B97 comparison, pairs such as B73 by CML247 required a much large number of traits to be correctly classified. This was likely due to these two varieties having very similar root systems. Interestingly, in cases where a large number of traits was needed, each individual trait contributed only a small amount to the classification. This can be seen in the small coefficient values for each trait. In cases where only a few traits are needed, those coefficients were much larger and dominated the model.

Lastly, the traits themselves fall into two categories, ones that were used over and over in classifying a large number of founders, and ones that were only used a handful of

times for few founder pairs. For example, while length distribution, maximum number of roots and major ellipse axis were key traits in multiple different models, maximum number of roots was only used while separating B73 from B97 or CML333, while width-to-depth ratio was only key in separating B73 from CML333. This is most likely because all of the models included B73, which has a small root system. The machine learning would consistently pick up on this feature when comparing it to founders with larger root systems, making traits that describe depth or size distinguish themselves over and over. Only when comparing B73 to other similar looking founders would the machine learning use other traits.

Analysis of temporal data

A great advantage of using the gel system is that the nondestructive imaging allowed us to image the same plant on different days during its growth cycle. The NAM founders were imaged on days 3, 6, 9 and 12 after planting allowing us to observe RSA changes over the course of that time period.

A majority of the traits showed a widening in the distributions of their values as the plants aged. For example, on day 3 the range of depths for all plants was from 2 to 6cm. By day 12, the depths ranged from 6 to 16 cm. The same was seen in a large number of other traits. This had the effect of making the different founders much more distinct from each other. Additionally the traits that showed the widest distributions were the

same traits that were repeatedly picked as the best classifiers during machine learning. This made sense as the model focused on the key differences between the varieties being tested, and traits that were especially variable made for the best predictors.

Interestingly, a few traits showed drastic changes at the early time points but stabilized by day 9 and 12. This was especially apparent with the solidity and average root width traits. On day 3, solidity had values in the ~.1 to ~.5 range, but by day 6 that range has dropped to ~.1 to ~.2. Following day 6 there was very little further change in solidity values among any of the founders. This was interesting, as for solidity to remain that constant any root growth within the confines of the root mass had to be balanced by outward growth. This suggests that the different founders maintain different roots densities, for at least the initial period of growth.

A similar pattern was seen in the average root width trait. On day 3, the widths ranged from 1.41 to 2.36, but then became progressively smaller on each successive imaging day (day 6, 0.85mm – 1.23mm, day 9, .75mm - 1.12mm, day 12mm, .75 - 1.05mm). Because of such a small amount of variability in the average root width, this made it a poor trait for differentiating most varieties. There were two exceptions. The founder CML228 had the second largest average root width, large enough that it showed up as the second best trait in the machine learning analysis. The founder Ki11 had the largest average root width (d03:2.01mm, d06:1.47mm, d09:1.34mm, d12:1.36mm). The difference

was so striking that this single variety could be easily picked out from all the others by eye alone. This was further seen in the machine learning analysis, as the average root width was by far the strongest trait for its classification.

It should be mentioned that the average root width trait is the average width of all the roots. That means that any information about the root widths of different root classes, for example primary vs lateral, is lost. Additionally this trait could also be confounded by the fact that the proportion of the different root classes vary between different founders. This is the primary reason for the overall decrease in average root width with time. The primary and seminal roots, both of which tend to be thicker, emerge first, resulting in larger average root width. Several days later the thinner lateral roots begin to emerge. Although shorter, they are much more numerous, and therefore drive down the average root width.

Comparison to other NAM founder datasets

Most of the NAM founder lines were selected to capture as much genetic diversity as possible. As a result the lines picked were either not intensively phenotyped, or the phenotyping was focused on aerial tissues. In two cases where the root systems of the NAM founders were examined, the phenotypes reported were not directly compared to the traits measured in our system.

Manavalan et al., 2011 examined the root length of the NAM varieties in a turface:sand mixture and reported depth as the length of roots laid out on a flat surface. This is different from the depth measured by our system, which is the distance from the top of the root system to the bottom of it irrespective of the path taken. To compare the two data sets we've manually measured the longest root in our plants, ranked them, and compared to the ranking obtained by Manavalan et al., 2011. The data sets were not correlated, likely due to the plants being of different ages as well as grown in different environments.

Clark et al., 2013 examined various root responses of the NAM founders to aluminum toxicity. The plants were grown in a pouch system, scanned on day 4, and analyzed using the RootReader2D software. Again comparisons of the primary root length of our plants to the control plants from Clark et al. 2013, did not show a positive correlation. The likely cause of this is the dissimilar time points and methods of phenotyping.

3D analysis

As previously mentioned, in addition to the 2D image phenotyping, our analysis pipeline has the additional capability to reconstruct 3D models from the 2D image sets. These in turn can be phenotyped to produce a separate set of 3D traits. Validation of the imaging platform with the ground truth model has shown that traits that attempt to

estimate 3D traits (Topp et al., 2013), such as volume and surface area, are much more accurate when calculated from 3D models then when estimated from 2D images. Unfortunately we were unable to use 3D phenotyping for this study. This was due to difficulties performing the reconstructions. A large number (between half and a third) of the plants produced faulty reconstructions. This included models missing roots, extremely thin roots, roots that split in two roots near the tips, roots that fused into a large unrecognizable "blob" or roots with shelf-like projections. A smaller fraction, about 1 in 10, was unable to be reconstructed at all. This is in contrast to rice image sets which could be reliably reconstructed most of the time, though the oldest roots almost inevitably had some root "blobs" where the concentration of the roots was the thickest.

Because the 3D reconstructions were so unreliable, and because the 2D traits alone were very efficient at separating different varieties, we found it prudent to focus our efforts on the 2D analysis. The reason for the poor reconstructions was likely a combination of ineffective software as well as too low of number of images in the image set. The two key reasons why the reconstruction software worked well for rice but not for corn are likely a combination of the size of the root systems and perspective distortion. The rice roots are smaller, and so the cameras are closer to the cylinder. Additionally a bit over half of the horizontal field of view of a rice image was empty as the cylinder only

occupied the central part of the image. On the other hand, the corn cameras are farther back and all of the horizontal field of view is occupied by the cylinder.

This in itself would not be a problem except for perspective distortion, which is a type of warping of an image due to the lens itself or the position of the camera relative to the object being photographed. This resulted in roots far away from the lens appearing smaller than they should be in relation to the roots closest to the lens. Additionally this made the roots appear to be on the wrong level, with the roots in the background appearing to be shorter than they should be. Lastly, each image in an image set was distorted slightly differently. Although some of this was addressed by the reconstruction method (Zheng et al., 2011), the large size of the maize root made the distortions especially problematic. Because of its smaller size and compactness, the rice root systems seemed to produce much more robust reconstructions.

A possible solution to this would be to use a lens that allows orthographic projection, such as a telecentric lens. Unfortunately these types of lenses are not only prohibitively expensive, but also there is currently no lens large enough to capture the whole field of view of a corn root system. This is because to produce a telecentric image, the size of the sensor has to be just as large as the object being photographed.

The perspective distortion is further exacerbated by the image density used. Although 40 images per set worked quite well for rice, in retrospect, this number seemed

insufficient for maize. This is especially apparent in cases where the roots have grown to the very edge of a jar. In cases like that, there was a large change in the position of the root in successive images. This tended to result in reconstructed models that either completely eliminated these far reaching roots or with roots that forked near their tips. Image sets with larger numbers of images, for example 80, could have potentially alleviated some of this, but would have also prohibitively increased the imaging time.

Chapter 4. Mapping of QTL root traits in the B73 x Ki3 mapping population

In Chapter 3 I surveyed 26 different varieties of maize, which showed a great amount of phenotypic variability in their root system architectures (RSA). These data were used to guide the selection process of a RIL family to be used for QTL mapping experiments. This chapter describes the selection process, as well as the subsequent data collection and QTL analysis.

Selection of mapping population

The NAM family is composed of 5000 RILs subdivided into 25 subpopulations (Yu et al., 2008; McMullen et al., 2009). Imaging the whole NAM population would be impossible with our current setup. Instead I selected a single subpopulation, B73 x Ki3, for QTL studies. Although the B73 x Ki3 subpopulation was selected, there is enough genetic variability between any of the lines that any of the subpopulations could have been a viable choice. I selected the B73 x Ki3 population for several different reasons. First I wanted the population whose founders were as phenotypically distinct as possible. Since all of the RIL subpopulations share the B73 parent, the goal was to pick the second parent to be as phenotypically distinct from B73 as possible. The RSA of B73 is actually remarkably average (Fig 12). Its roots are not the deepest, but also not the shallowest. It has a fairly compact root mass, but there are varieties that are more compact and others

that are much less so. This allowed me to narrow down the pool of candidates to ones with small and compact RSA, such as HP301, or ones with large and wide ranging root systems, such as Ki3 or Tx303. Because B73 was closer in RSA to the small types and because the large varieties tended to look healthier than the lines with smaller RSA I decided to select a founder with a large and extensive root system. I selected the Ki3 founder since it has the second most extensive root system of all the founders. Although the Tx303 founder has an even larger root system, I decided against it due to possible space constraints. The Tx303 founder has a very extensive root system and tends to run out of space extremely rapidly. This would be especially severe if transgressive segregation would produce RILs with RSAs even larger than the Tx303 founder (Rick and Smith, 1953).

I also wanted to select genetically distinct founders. This has already been accomplished during the creation of the NAM population as the 26 founders used were selected to encompass as much genetic diversity as possible (Flint-Garcia et al., 2005; Liu et al., 2003). Furthermore, the B73 and Ki3 founders are also historically very distinct. B73 belongs to the stiff stalk heterotic group and was developed in Iowa in 1972 (Smith et al., 2004), while Ki3 belongs to the tropical heterotic group and was developed in 1984 in Thailand (Udomprasert et al., 1997). This made these two founders both especially genetically and phenotypically distinct, making this a good choice for QTL mapping.

Experimental design

The B73 x Ki3 population is composed of 200 individual plant lines. Of these I was able to obtain seeds for 193, of which I was able to image 186. A few of the lines showed poor germination and/or were unable to grow in the gel system. Unfortunately marker data was unavailable for an additional 11 lines, bringing the final count to 175 lines. Although ideally the full population would have been used, 175 RIL lines were more than sufficient to proceed with QTL mapping. Each plant was imaged on days 4, 6 and 8 after planting, and at least 3 replicates for each RIL was obtained. This was different from the days used for imaging the NAM founders due to the large size of RSA of some of the RILs. Plants were grown in batches of around 30 plants. Ki3 and B73 varieties were grown as controls although some were lost due to contamination or poor growth after transplantation to gel. All of the plants were grown in one of three identical growth chambers (T6, T7, T8) in the Duke Phytotron. Plants were grown on a 16/8 hour day/night cycle, 28ºC day and 25ºC night with 800 micromol PAR $m^{-2}s^{-1}$ light. Additionally the different replicates for each line were grown at different times or different growth chambers. All other protocols were used as described in Chapter 2.

Distribution of trait values across the mapping population

Prior to analyzing the data, we wanted to look at the trait distribution. In cases where the traits being phenotyped are quantitative, a normal distribution could have been

seen. Additionally, it's likely that some of the individual lines would demonstrate transgressive segregation and have trait values lower or higher than the average values of the population founders. Indeed, we found this to be the case for most of the traits. Figure 14 shows a histogram for all 17 traits for all imaging days. Additionally the values for the parent varieties are shown as horizontal lines.

Most of the traits (11 of 17), show the kind of distribution exemplified by network area. The average trait values for the parents fall within, but not at the very edge, of the tails of the distributions, indicating that transgressive segregation is occurring. The distributions are also shifted progressively to the right with older datasets. Again this is expected as this indicates growth of the root over the course of the three imaging days. Additionally the distributions become broader, while the peaks became shallower, which indicates that there was more variability between RILs in the older day data sets. In the early stages of growth, the RSA seem to be more similar to each other, but, with time, they developed much more distinct RSA. This likely contributes to larger heritability values in the older day dataset, as discussed in the next section. Overall, these data match what would be expected of any RIL population.

A smaller subset of the traits (5 of 17), such as solidity, length distribution, bushiness or width-depth ratio showed a different pattern (Fig 14). The distributions themselves remained somewhat static. This was because all of these traits are ratios of two

different traits. Because the component traits increased proportionally to each other, the ratio of the two remained the same. The fact that the proportionality remained similar over time suggested that, despite a large amount of growth, some of the aspects of the overall architecture, such as root density or thoroughness of exploration, of these lines remained relatively constant.

The last trait, average root width, was similar in almost all respects to the first group, except that the distributions for the older days shifted to the left while the distributions themselves became narrower. Furthermore the shift from day 4 to 6 was more pronounced than the shift from day 6 to day 8. Although unexpected, this could be easily explained. The root system of the early seedling was made up of the seed, a short primary and several short seminal roots. This artificially inflated the value for average root width. As the plant grew, the contribution of the seed to the overall average root width decreased. Additionally lateral roots begin to emerge over time, which were thinner than the primary, crown and seminal roots. The larger shift from day 4 to 6 was likely due to the small ratio of root to seed width at the early stages. By day 8 the seed was a negligible portion of the total and the further leftward shift was likely due to the emergence of thinner lateral roots. The larger thickness of the maize versus rice roots gave us much higher contrast with the background allowing us to capture and include more roots in the analysis.

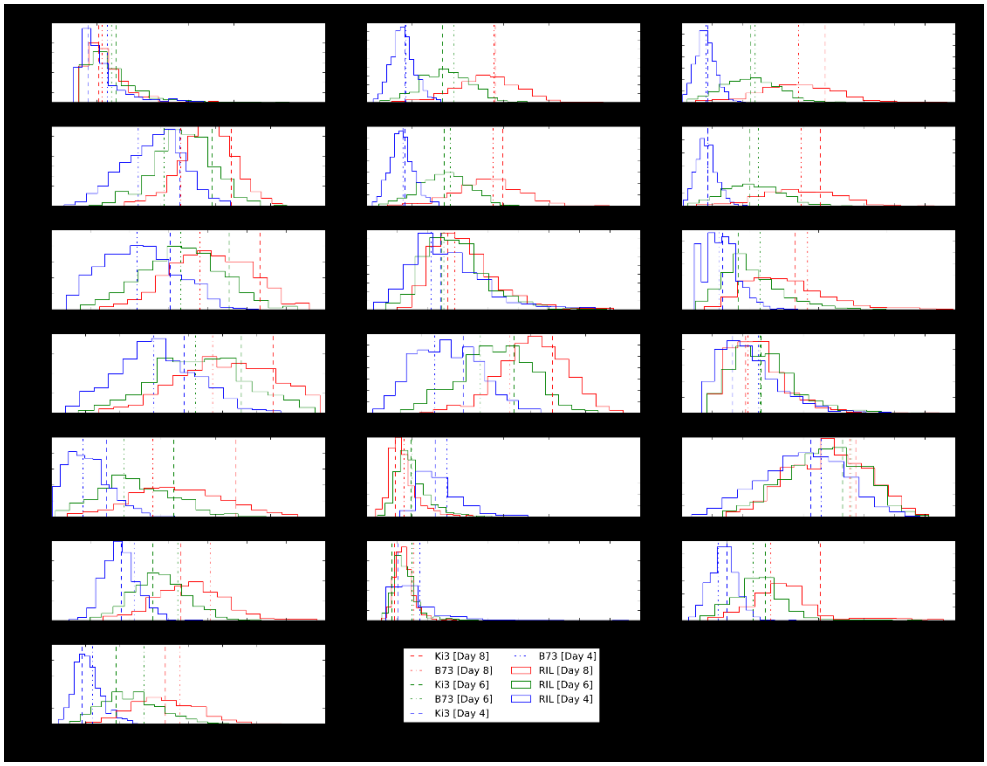


Figure 14: Trait distribution values for the B73 x Ki3 RIL population. Horizontal lines indicate average values for the B73 and Ki3 parents.

Broad sense heritability among traits

Prior to the QTL analysis the data were first analyzed using JMP Pro 11 (SAS Institue Inc, 2014) to determinate heritabilities and to see if there was an effect of plants being grown in different chambers. This was done using random effect model ANOVA for both the RIL families as well as the chamber used (Fry, 1992). Broad sense heritability (Fig 15) is the proportion of the genotypic variance (RIL families) over the total phenotypic variance. Most of the traits were highly heritable, especially with the older plants, with most traits having heritabilities over 30%, and a few reaching 50%. This meant that fewer samples would be required for QTL mapping, as most of the variability between the different RIL lines is due to genetic differences and not environmental factors. It's especially encouraging that the traits that define the differences between the B73 and Ki3 varieties, such as depth, network convex area, solidity and minor ellipse axes, have especially high heritabilities. In addition to heritability, the ANOVA also indicated that in a few cases there was a chamber effect. This effect was minimal and barely statistically significant. To eliminate this effect, the least square means provided by the ANOVA were used in further analysis as this factors in the chamber effect and eliminates it.

Trait	Day 4	Day 6	Day 8	#UniQTL	Included in multiQTL
Median Number of Roots	18.85	13.47	22.16	2	
Volume	34.13	31.51	16.67	2	
Bushiness	8.11	6.80	15.66	3	
Max Width	30.53	34.14	36.65	3	x
Length Distribution	13.00	13.46	26.86	3	
Ellipses Axes Aspect Ratio	19.70	29.89	48.78	3	
Specific Root Length	32.31	33.63	36.93	3	
Perimeter	31.12	33.27	44.13	4	x
Surface Area	30.54	31.25	31.77	5	
Network Area	30.94	32.38	34.80	6	x
Total Length	29.45	32.82	43.10	6	x
Width-Depth Ratio	21.79	28.80	40.49	6	x
Maximum Number Roots	17.60	30.21	40.84	6	x
Average Root Width	27.59	38.69	49.84	7	x
Solidty	30.62	43.10	43.21	7	x
Minor Ellipse Axes	32.22	44.27	54.99	8	x
Depth	29.55	42.22	49.85	8	x
Network Conex Area	31.81	42.64	49.79	9	x
Major Ellipse Axes	29.86	41.57	43.75	12	x

Figure 15: Broad sense heritability for the B73 x Ki3 RIL population. The range of values have been divided into thirds. Values in red fall in the lower 1/3, values in yellow fall in the middle 1/3, and values in blue fall in the top 1/3.

Quantitative Trait Locus (QTL) mapping

The QTL analysis was performed using the Linux command line version of the QTL Cartographer package ver 1.17f (Basten et al., 1994, 2004). This involved creating the required input files, followed by sequentially running several different applications. What follows is a brief description of the process as the majority of the detail is covered in the QTL Cartographer manual (Basten et al., 2004).

Prior to running the analysis, the raw data was formatted to the QTL Cartographer format. The standards required are provided by the manual and involved creating two

input files, one containing the linkage map, and a second containing the marker identity for each RIL as well as the average trait values for each RIL. These two files were converted to QTL Cartographer format using the Rmap and Rcross programs provided by the QTL Cartographer package.

The QTL mapping was first performed using SRmapqtl. This program used stepwise regression to search for QTLs. This technique ranks all of the markers based on their effect on the quantitative traits (Basten et al., 2004). This can be done in one of several ways. Forward addition works by successively adding the best marker to the model until the model cannot be improved further, while backward elimination does the opposite by successively removing worst markers. In both cases an F-statistic is used to determine significance. The last method, and the one used here, combines both of the above. The markers are added as in the forward addition model, but after each step, the already added markers were re-checked and eliminated if no longer significant. A separate significance cutoff was used for the addition and removal. The end result of this mapping was a list of ranked most significant markers, with the most significant markers being the most likely location of the QTL. However this technique has a severe flaw in that its model assumes a single QTL for each trait, and therefore the presence of secondary QTLs can have interfering effects on the mapping.

A second method called Composite Interval Mapping (CIM), assumes multiple QTL locations (Zeng, 1993, 1994). It works by selecting a subset of markers as loci to serve as covariates in the model. During the analysis these markers are used to temporarily mask out neighboring regions of the genome to eliminate their potential effect on the current area being examined. Obviously, selection of the correct markers as the covariates is a key step. If the markers are selected correctly, so that they truly reflect locations of neighboring QTLs, than the accuracy of the CIM will be improved. Since rough mapping was already performed using stepwise regression, the top ranked markers generated by that method were used as the covariates for the CIM mapping.

The CIM mapping itself was done using the Zmapqtl program using model 6 and including at most 5 markers from SR mapping as covariates. One centiMorgan (cM) walking distance was used since the linkage map used had the same marker density. The analysis was run twice, a single time without permutation to perform the real analysis, and 1000 times with permutations to establish a significance threshold. The permutation analysis randomizes the linkages between the genotypic and phenotypic data and saves the largest likelihood ratio (LR) for each trait for each iteration. An appropriate percentile value, in this case 95% and 99%, can then be used as the significance threshold for the results from the non-permuted analysis.

Calculation of confidence interval

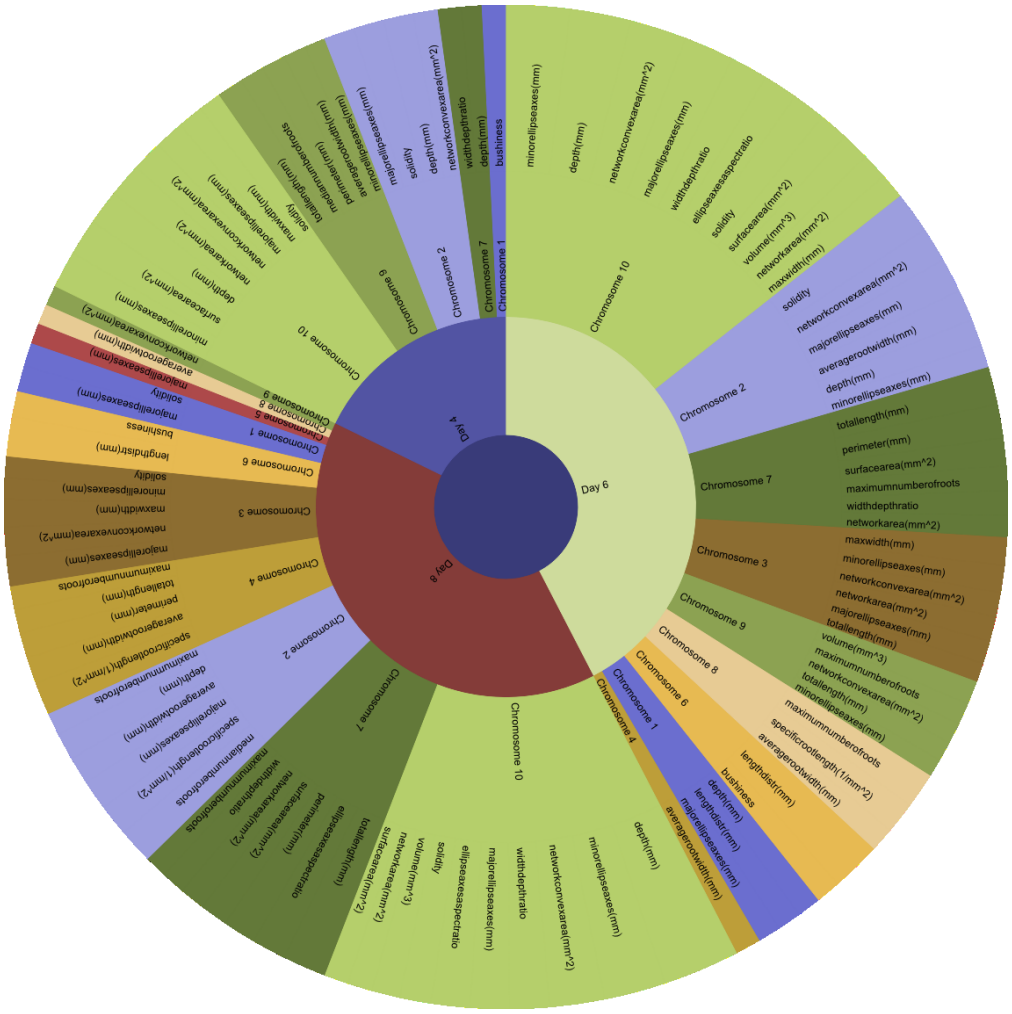
Due to the nature of the MLE, the marker with the highest LR is most likely the closest to the location of the causative gene. Confidence intervals (CI) are generally constructed to demarcate a likely region of the chromosome for the location of the QTL. The most common approach is to use 1- and 2-LOD CI (Lander and Botstein, 1989). This is done by selecting the marker with the strongest LOD score x, and finding the closest neighboring marker where the value passes either x – 1 LOD or x – 2 LOD threshold. Alternative approaches involve bootstrapping (Visscher et al., 1996) or analytically derived formula (Weller and Soller, 2004), but none of these approaches are ideal. In particular LOD cutoff methods tend to generate CI that are too small and can be highly affected by the effect size of the QTL (for example, a theoretical QTL with LOD score 40 would have the 2-LOD cutoff of 30.79, while most LOD scores tend to be considered significant at LOD scores above 3.5). Here we show CI in two ways, using 1- and 2-LOD cutoffs and using permutation threshold cutoffs. The LOD cutoffs were calculated by starting with the strongest marker and scanning outward until the appropriate thresholds were met. In order to allow for fluctuation in the LOD score values, a gap size of 5 markers was allowed. Lastly the final cutoff points were interpolated between the threshold marker and the next subsequent marker.

The permutation based CI were calculated similarly, the only exception being that the cutoff points were selected based on the permutation analysis. For each trait the outer CI limit was set to the alpha = 0.05 significance threshold, while the inner CI limit was set to the alpha = 0.01 significance threshold. As expected, in cases of QTLs with large effects, the LOD CI produced a smaller CI as it tended to not include some of the significant markers, while the permutation based CI did since it included all significant markers irrespective of the 1- or 2- LOD significance cutoff. In cases of QTL with large effects, the LOD CI tended to produce larger CI since it would include markers that did not meet the significance threshold, while the permutation based CI would eliminate those. For example the strongest QTL had 50.52-55.53-58.53-60.36, 9.84cM spanning, LOD CI, and a 46.16-46.28-66.23-66.36, 20.2cM spanning, permutation based CI. On the other hand the weakest significant QTL had 47.51-51.79-57.01-62.23, 14.58 cM spanning, LOD CI, but had only a single significant marker at the alpha 0.05 permutation significance level.

The CI for the QTLs reported here (Fig 17) have been assigned based on the 1- and 2-LOD score method, with a gap of at most 5 markers allowed. I did this because the high density of markers tended to inflate the number of QTLs by breaking up a region that contained a single non-significant marker.

Distribution of QTLs

For the 57 traits mapped (19 traits on 3 different days), we located a total of 102 QTLs. 69 of these were significant at alpha = 0.01, while the remaining 33 were significant at alpha = 0.05. The QTLs were unevenly distributed among the days. There were 20 QTLs for day 4 traits, 42 QTLs for day 6 traits and 41 for day 8 traits. The smaller number of day 4 traits was likely due to less apparent variability in the population at that young age. The effect size of QTLs ranged from 5.48% to 23.8%, which was higher than expected based on previously mapped flowering QTLs mapped in the whole NAM population (Buckler et al., 2009).


Day 4
Day 6
Day 8
Chromosome 1
Chromosome 2
Chromosome 3
Chromosome 4
Chromosome 5
Chromosome 6
Chromosome 7
Chromosome 8
Chromosome 9
Chromosome 10
bushiness
depth(mm)
widthdepthratio
networkconvexarea(mm^2)
minorellipseaxes(mm)
majorellipseaxes(mm)
solidity
networkarea(mm^2)
maxwidth(mm)
totallength(mm)
perimeter(mm)
surfacearea(mm^2)
maximumnumberofroots
networkconvexarea(mm^2)
averagerootwidth(mm)
specificrootlength(1/mm^2)
lengthdist(mm)
volume(mm^3)
ellipseaxesspectratio
surfacearea(mm^2)
averagediameter(mm)
numberoftips
networkwidth(mm)
networklength(mm)
networkarea(mm^2)
maxdepth(mm)
networkconvexarearatio
networkaxesratio
networkperimeter(mm)
networkdiameter(mm)
networkdepth(mm)
networkwidthdepthratio
networkmaxdiameter(mm)
networkmindiameter(mm)
networkmaxdepth(mm)
networkmaxwidth(mm)
networkmaxarea(mm^2)
networkmaxperimeter(mm)
networkmaxconvexarea(mm^2)
networkmaxellipseaxes(mm)
networkmaxaxesratio(mm)
networkmaxdiameterofroots
networkmaxdepthofroots
networkmaxwidthofroots
networkmaxnumberofroots
networkmaxsurfacear

(Fig 16). With a few exceptions, the QTL tended to cluster together. There were only 2 single QTL across all of the chromosomes, but 9 multi-QTL, determined based on overlapping CI, clusters. With a few exceptions, the QTLs tended to cluster together. There were only 2 single QTLs across all of the chromosomes, but 9 multi-QTLs, based on overlapping CI clusters. The largest cluster was located on chromosome 10 and included 25 of the 29 QTLs on that chromosome. Eight of the 25 ranked as the top most significant and largest effect QTLs. The remaining 4 QTLs made up a 2nd cluster which was located in close proximity to the large cluster, suggesting that those QTLs could potentially also be a part of the same cluster. The large cluster on this chromosome, as well as the location of many strong QTLs in it, makes this the most promising candidate for future work. The second most promising chromosome, 2, had a major cluster composed of 14 QTLs, and a minor, but neighboring, cluster composed of 2 QTLs. Next, chromosome 3 had a single cluster that included all 11 QTLs on that chromosome. Chromosome 7 had a major cluster composed of 10 QTLs and a minor cluster composed of 5 QTLs. Chromosome 9 had a major cluster composed of 9 QTLs and a minor neighboring cluster composed of 2 QTLs. The remaining chromosomes had smaller clusters composed of 6 or fewer QTLs. Although the clusters are smaller they should not be discounted. For example chromosome 4 had a single cluster composed of 6 QTLs, but 5 of those meet the alpha 0.01 significance threshold and effect sizes that range between 7.1 to 10.1%. This could indicate

that while the large clusters, such as on chromosome 10, have pleiotropic effects on RSA, some of the smaller clusters could have much more narrowly focused effects on only certain aspects of the RSA.

QTLs for the same trait, but different imaging day also tended to be found together. For example on chromosome 10, depth, major ellipse axes, minor ellipse axes, network convex area and solidity all have QTLs for all 3 imaging days within overlapping CI. Width-to-depth ratio, volume, surface area, network area, max width and ellipse axes aspect ratio have 2 of the 3 days within overlapping CI. Additionally the last imaging day for both surface area and network area have a QTL located within several centiMorgans for the other two days. The same kind of stacking can be seen on chromosomes, and with a few exceptions, most of the traits are found overlapping at least one other, but different day, trait.

The effect size of the QTLs themselves ranged from 5.5% to 23.8%. This was in contrast to flowering time which was found to be controlled by many small effect QTLs. The strongest effect QTLs were for depth (day 8: 23.8%, day 6: 18.85%) and minor ellipse axis (day6: 19.8% day 8: 19.1%), both of which were distinguishing traits for the founders of this population. The depth QTLs was located on chromosome 10 and had an additive effect of -7.64 mm for day 8 and -6.01 mm for days 6 (that is the B73 allele results in shallower roots). The minor ellipse axis QTLs were located in the same cluster and had

additive effects of -4.93 mm for day 6 and -5.82 mm. All of the QTLs found are reported in Appendix C.

A lot of the QTLs had additive effects suggesting the larger root phenotype of the Ki3 founder. This isn't surprising as the Ki3 founder had larger values for most traits for the founders. Traits such as depth, minor ellipse axes or network convex area all had larger values for the Ki3 founder since it had a larger root system. On the other hand, traits that had larger values for the B73 founder are also traits for which the QTL additive effects are due to the B73 allele. The prime example of these are length distribution (B73 had more roots close to the surface then Ki3), solidity, bushiness and width-to-depth ratio.

This can also be seen in the context of clusters. For example, most of the QTLs that go into the major cluster on chromosome 10 indicate that the additive effect support Ki3-like phenotypes with the exception of solidity and width-to-depth ratio. This is very similar to the tradeoffs described in chapter 3. It suggests that a single gene in that particular region controls a large portion of the RSA, and due to the fact that it is a single locus, only certain root archetypes are possible. The QTL cluster on chromosome 9 shows the opposite effect. Although no significant QTLs for solidity were found in that cluster, the additive effect of markers for the solidity traits in that area suggest a smaller, B73-like RSA, while a majority of the additive effect of the other traits are due to Ki3 alleles. The same pattern can be seen in the major QTL cluster on chromosome 2.

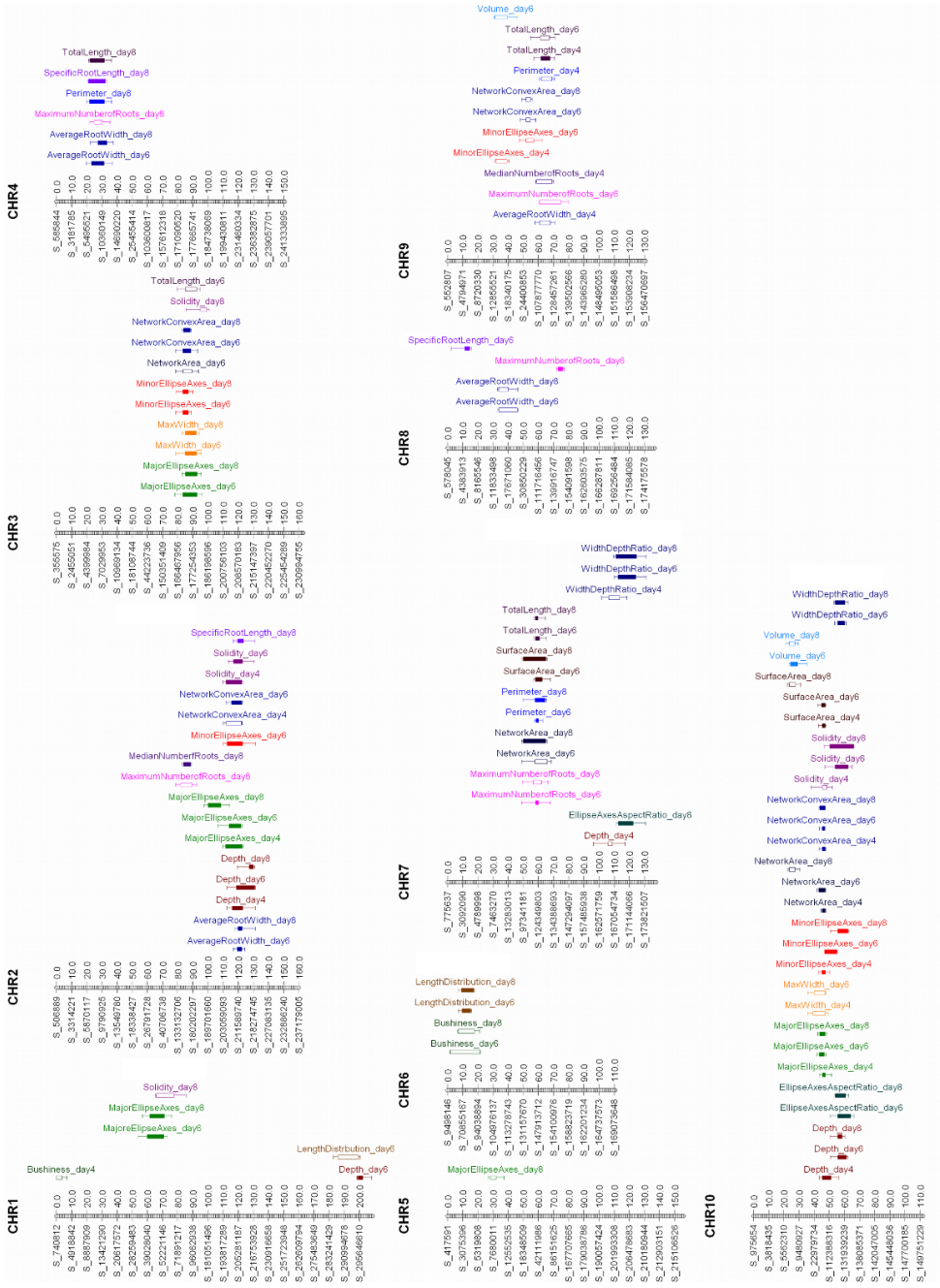


Figure 17: Local of all QTL across the maize chromosomes. Solid boxes indicate significant at alpha = 0.01. Hollow boxes indicate significant of alpha = 0.05. The outer whiskers of each box indicated the 2-LOD CI. The outer boundary of the boxes indicated the 1-LOD CI.

Chapter 5. Discussion and Future Direction

Discussion

I've described the development and adaptation of a gellan gum based system for imaging of maize root systems, as well as the computational pipeline for processing and phenotyping of the resulting images. In general, phenotyping systems present tradeoffs between the throughput of sample collection and the level of detail captured. The gel system described in Chapter 2 has the advantage of allowing both (Iyer-Pascuzzi et al., 2010; Topp et al., 2013). We were able to produce very accurate images of whole root systems, preserving their 3D architecture, and also do so at a high enough rate to allow for surveying large numbers of maize varieties and QTL mapping. Additionally this system uses a standardized media which has the advantage of controling for environmental effects, and is nondestructive, allowing for repeated measures of the same root system at different times during growth.

We used this system to perform a survey of RSA of the 26 founder lines of the NAM population (Yu et al., 2008). These constitute a large portion of the total genetic diversity in maize (Buckler et al., 2009). Not surprisingly we observed a tremendous amount of phenotypic variation between the different founders, bound on the extremes by two distinct ideotypes. One, exemplified by Ki3, had a much wider and deeper RSA that tended to explore and encompass a large amount of space, but left behind a large

amount of empty space, which resulted in pockets of space surrounded by individual roots. This type of RSA could have several benefits. The expansive root system could serve as a type of a scout network, allowing the plants to find the best location before fully allocating limited biomass. Once a particularly nutrient rich area was found, the plant could preferentially develop further lateral roots there and use the previously established primary and seminal roots as highways to shuttle nutrients to the rest of the plant (Jobbágy and Jackson, 2001). Alternatively, since this type of RSA would result in much deeper root systems, plants possessing it could potentially be much more proficient at acquiring deep subsurface water (Lynch, 1995). This would make the deep and exploratory RSA especially beneficial to plants grown in arid and drought-prone locations.

At the other end of the spectrum were varieties such as B73, with small and compact root architectures. These tended to produce dense root networks close to the site of germination and then slowly expand outwards. Even then, they would never reach the size of the large exploratory varieties. This type of RSA could have its own benefits, and seems especially tailored for agricultural use. The distributions of roots toward the top of the soil would allow these plants much easier access to fertilizers, which tend to stay in the top soil (Jobbágy and Jackson, 2001), while the compact root mass would allow for increased planting density (Hammer et al., 2009).

Between these extremes were varieties that traded off between the expansionist and compact ideotypes. For example the founder OH7B had a root system that was shallower, but also wider, then B73. No single variety had a root system that was both dense and expansive. This reinforces the idea that there are physiological constraints on the distributions of roots and that different varieties prioritize allocation of biomass to different aspects of their RSA.

To locate the genetic factors that control root architecture we imaged and phenotyped the B73 x Ki3 mapping population. We found very high heritabilities for a large portion of the traits (Fig 17), which further illustrated the power of the gellan gum based system. Due to the reductionist nature of the system, a large portion of the environmental effects was eliminated, and allowed us to focus on the phenotypic differences due to the different genetic background of the B73 and Ki3 parent lines.

Next we used these data to conduct a QTL mapping analysis. We located a combined 102 QTLs across 17 traits and three time points. A majority of these were grouped into 9 separate clusters, with an average of one cluster per chromosome. Additionally in most cases QTLs for different days, but the same trait, also grouped together. This was identical to what we was previously seen in QTL mapping experiments using rice (Topp et al., 2013). This suggests that many different aspects of RSA are controlled by several key loci. The effect size of the individual QTLs ranged from 23.8%

to 5.49%. These were larger than the previously mapped flowering QTLs in this population (Buckler et al., 2009). This could be in part due to different levels of genetic control for root traits versus flowering time traits, but could also be due to different ages of the plants. Most interestingly the tradeoffs we observed between types of RSA among the NAM founder lines were also seen in the QTL clusters. The QTLs in each cluster could be divided into those that had additive effects due to either the Ki3 or B73 parent. The traits themselves could be divided into those that supported either an exploratory architecture, such as depth, or those that supported a compact architecture, such as a large solidity. With the exception of one cluster, all clusters had some QTLs with additive effects from both parent, and which trait was linked to which QTL went hand in hand with the type of architecture that trait would describe. For example the cluster on chromosome 10 had large additive effects from the Ki3 parent for the traits supporting expansive architecture and large additive effects from the B73 parent for the traits that supported the compact architectures. The opposite was true for the cluster on chromosome 9. This suggests a strong, almost inseparable, linkage between the two types of architectures and suggests that having both an expansive and dense root system would be unlikely. The one exception, the cluster on chromosome 6, hints otherwise. This cluster was composed of only 4 QTLs, but the additive effect of all were due to the B73 alleles, suggesting that this

cluster could potentially result in a more dense root system irrespective of the how wide spread the roots would be.

Future Directions

This work has located several clusters of QTLs for root traits in maize. The most interesting of these were the clusters composed of large numbers of QTLs, such as those on chromosome 2, 3, 7, 9 and 10. These would be ideal candidates for further fine mapping experiments. This could be accomplished through several approaches. One would be to generate near-isogenic lines (NILs) through the use of residual heterozygosity and heterologous inbred families (HIFs) (Tuinstra et al., 1997). This would allow us to isolate and further partition the fragment of the genome under the cluster and allow for much finer mapping. This would be greatly aided by the large number of genetic markers already available in the NAM population.

Additionally comparative approaches could aid in narrowing down the possible list of candidates. These include comparing RNA expression levels (Hansey et al., 2012), comparison to previously mapping maize traits and meta-QTL analysis (Tuberosa et al., 2003) or comparisons of gene lists under our QTL clusters to those of orthologous genes under QTLs in other plants (Lyons et al., 2008).

Alternatively due to the large number of markers already available, the clusters could be used for marker assisted breeding. Although at this point we do not have enough

information to select for individual traits, the information we presented here could be used for breading for specific types of RSA as a whole. The chromosome 10 cluster could be especially useful for breeding for deep and expansive architectures. The obvious first step of this process would be validate our results in soil experiments and in adult plants.

In summary our gel phenotyping system has allowed us to uncover the rich variability in root system architecture of maize. Furthermore, we demonstrated functional tradeoffs between different root architectures and have begun the process of locating the genes that control it.

Appendix A. Pipeline Components

Pipeline framework (Including Review, Export, Filter, etc) - Provides a graphical interface for all other components. Originally written by Bradley Moore. Further improvements and maintenance by Vladimir Popov.

Scaling and cropping - Part of the framework. Allows setting of scale and cropping of raw images. Written by Bradley Moore. Further improvements by Vladimir Popov.

Recrop - Part of the framework. Allows further cropping of already cropped images. Written by Vladimir Popov.

Gia2D - Functionally provided by Gia Roots. Allows processing of cropped or recropped images. Creates thresholded images and calculates 2D traits. Written by Taras Galkovskyi, Yuri Mileyko and Joshua Weitz Integrated into pipeline by Bradley Moore and Vladimir Popov

Rootwork - Functionality provided by Rootwork Allows reconstructions of 3D models from 2D thresholded images. Written by Ying Zheng. Modified for and integrated into pipeline by Vladimir Popov.

Skel3D - Functionality provided by Matlab script. Allows for calculation of skeleton based 3D traits. Written by Joshua Weitz. Integrated into pipeline by Vladimir Popov.

Gia3D_V2 (Legacy version: Gia3D) - Functionality provided by Gai Roots Allows for calculation of 3D traits from 3D reconstructions Written by Joshua Weitz. Integrated into pipeline by Vladimir Popov

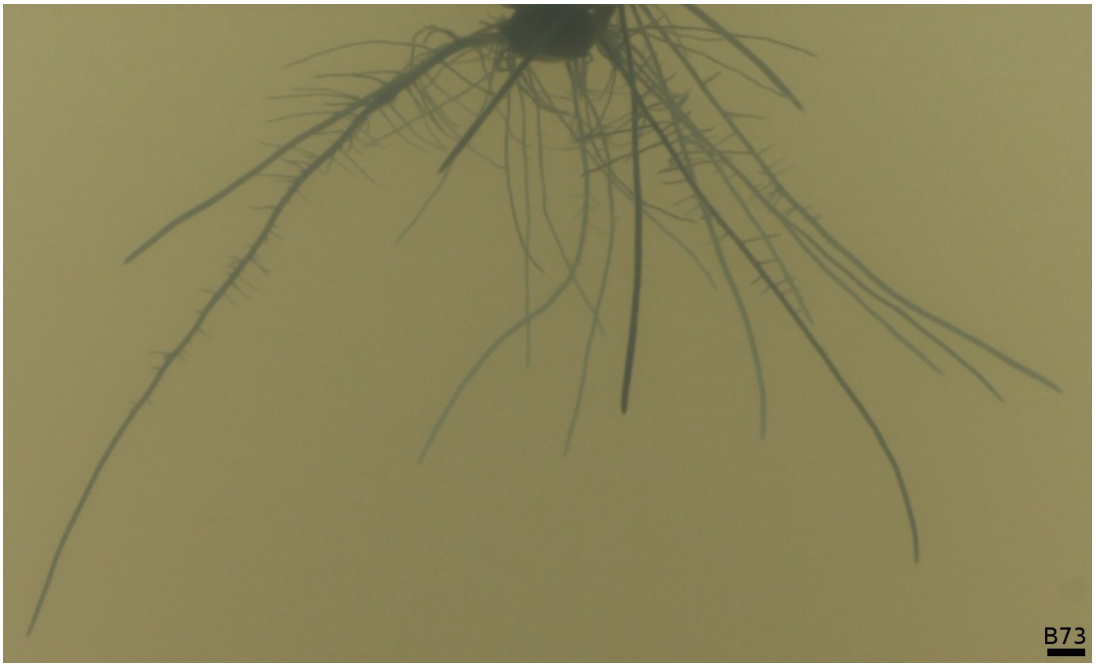
QC - Functionality provided by a Python script. Allows for creations of arrayed mosaic images (cropped, thresholded, or skeleton) composed from whole image sets. Written by Paul Zurek. Integrated into pipeline by Vladimir Popov.

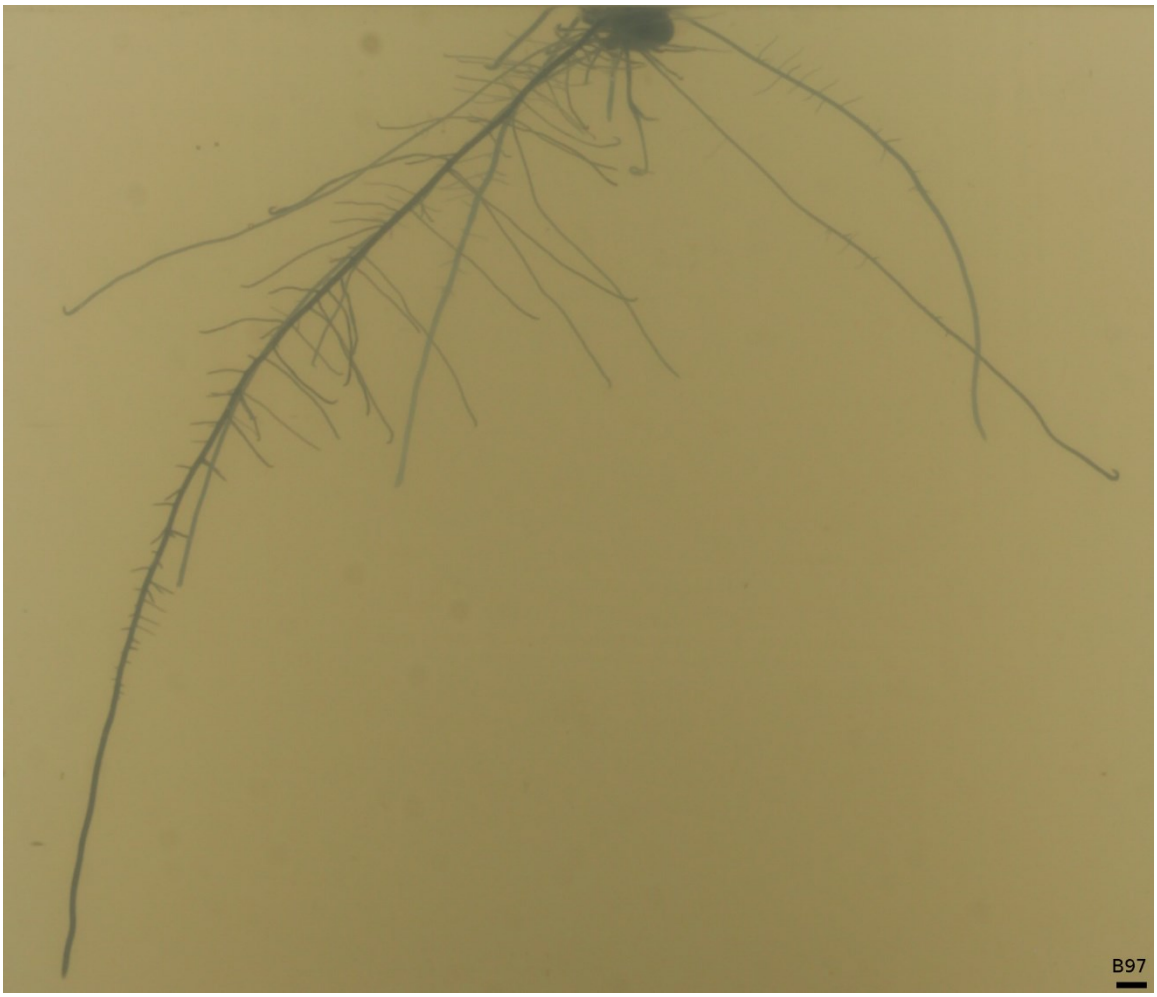
QC2 - Functionality provided by a Python script. Allow for quality control of thresholding. Written by Paul Zurek. Integrated into pipeline by Vladimir Popov.

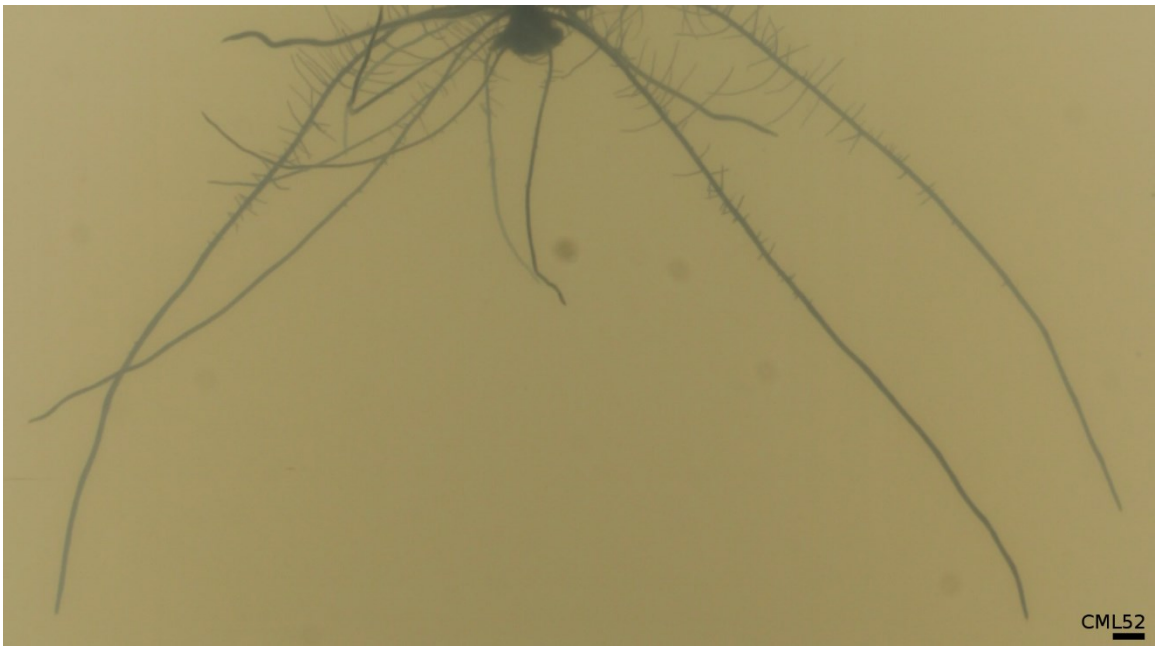
QC3 - Functionality provided by a Python script. Allows for quality control of 3D reconstructions. Written by Paul Zurek. Integrated into pipeline by Vladimir Popov.

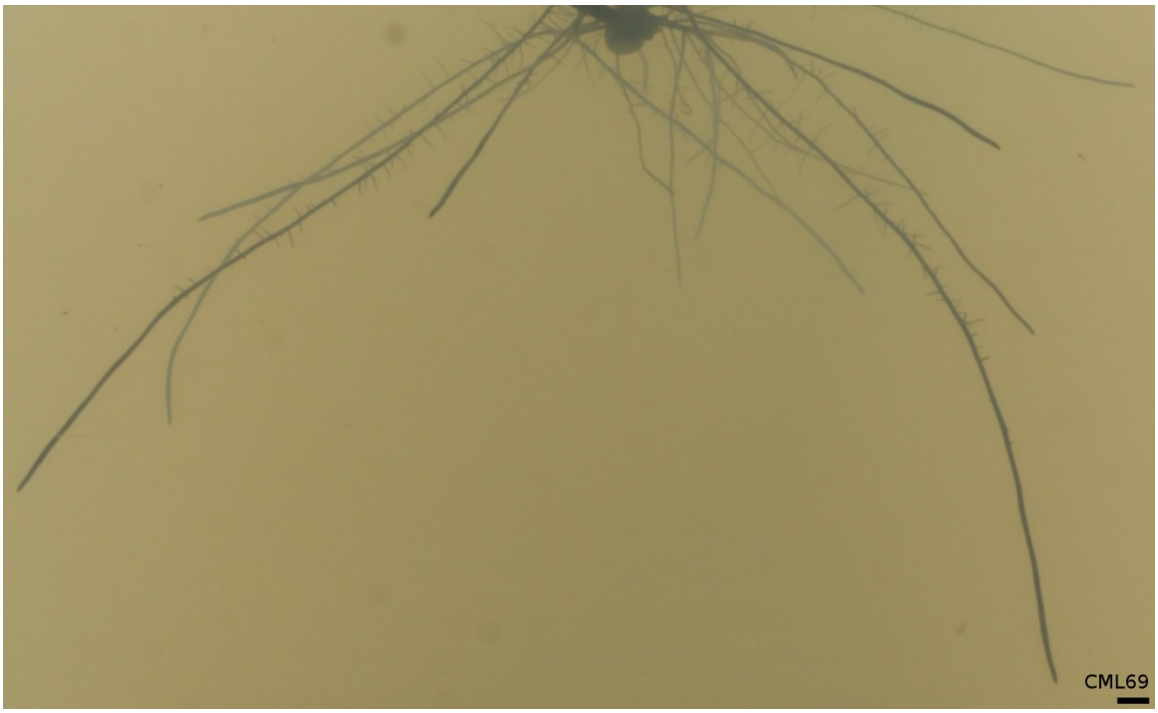
Appendix B. NAM founder lines

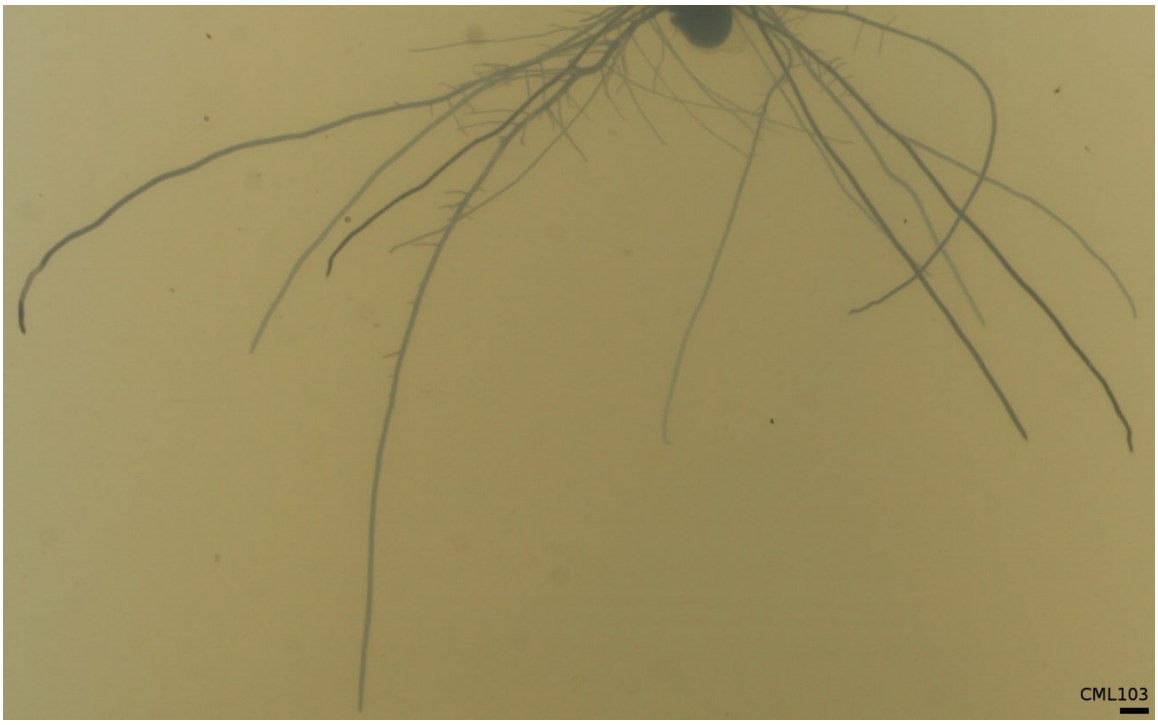
A representative sample for all founders at 9 days after transplantation. The black bar represents a scale of 5 centimeters.

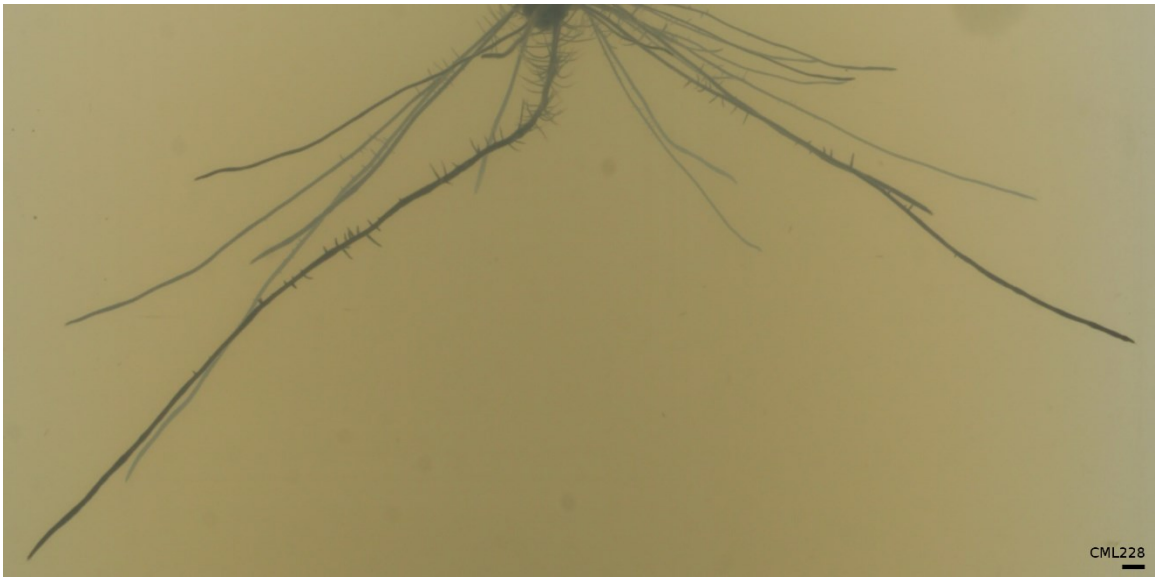


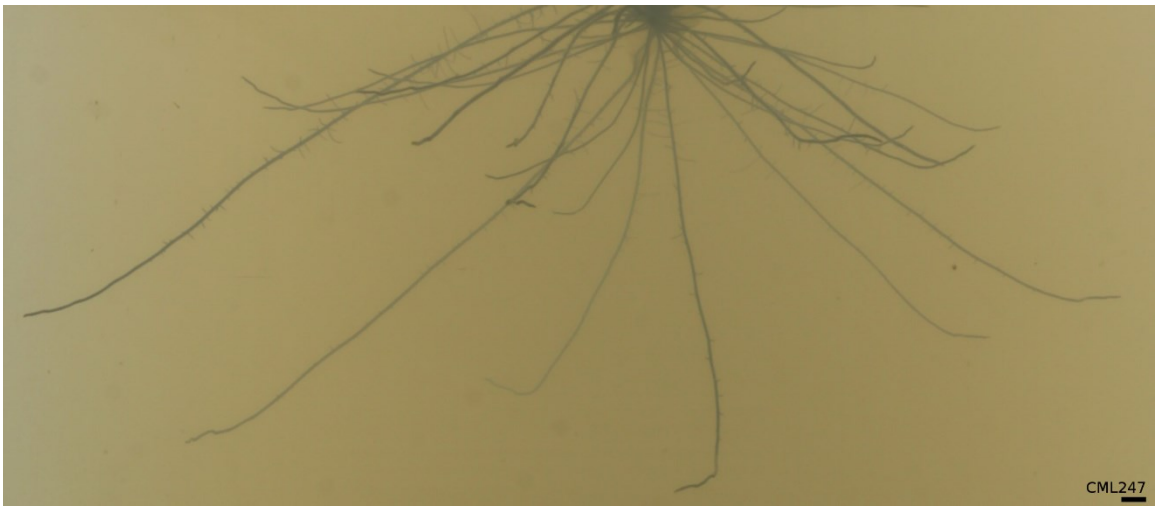

B97

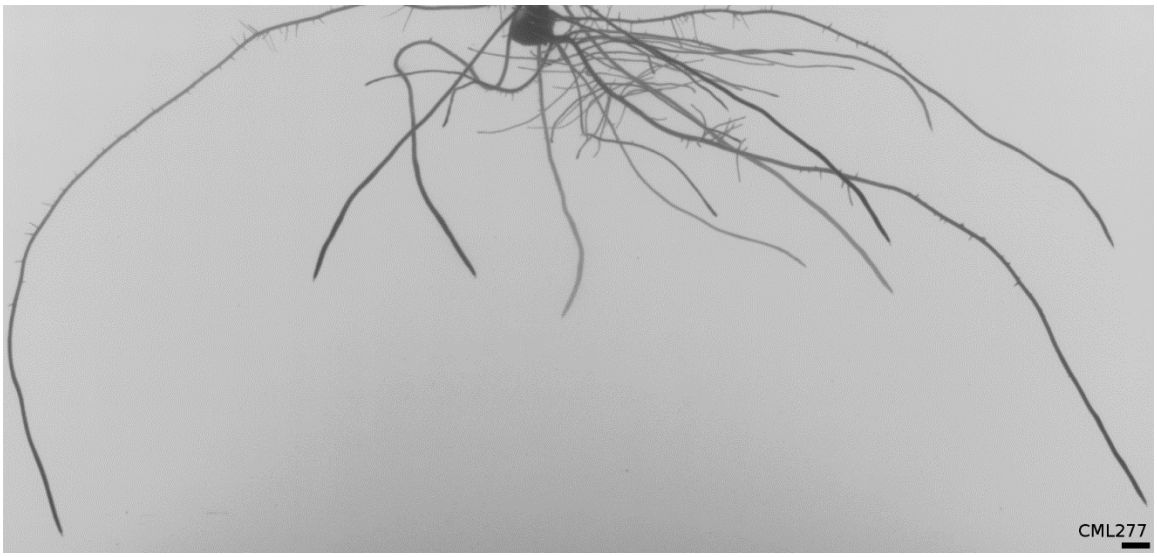

CML52

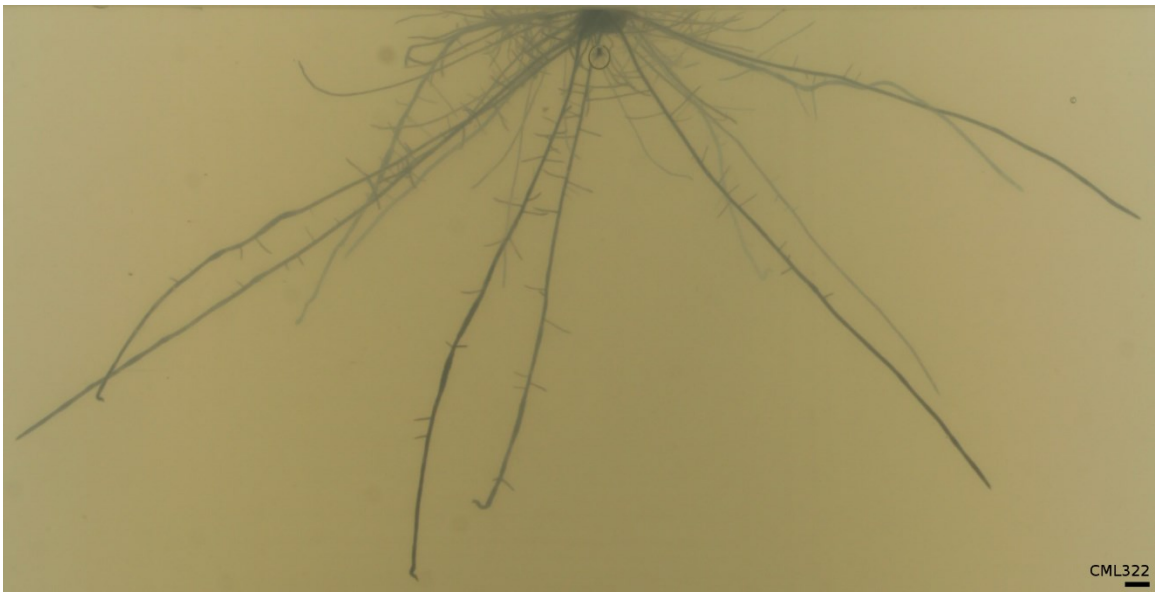

CML69

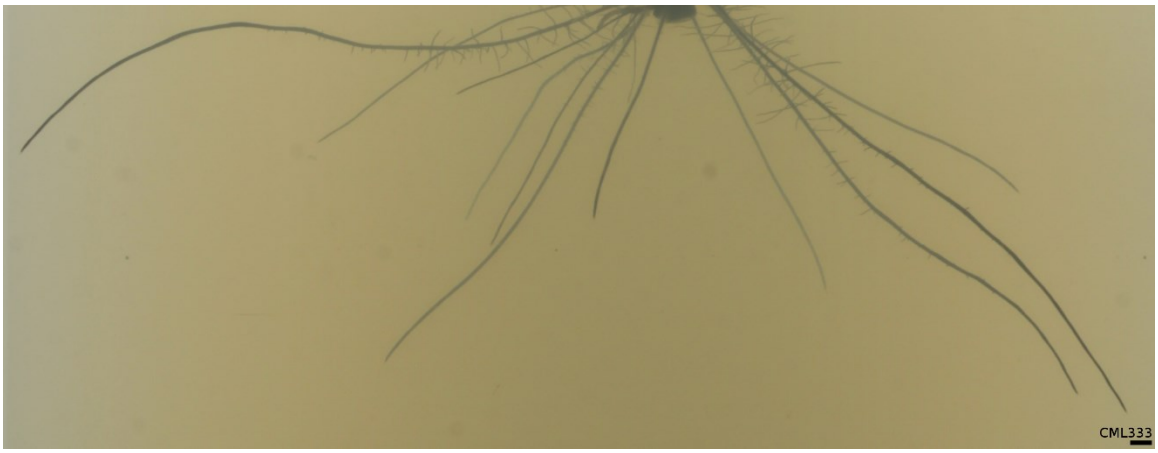

CML103

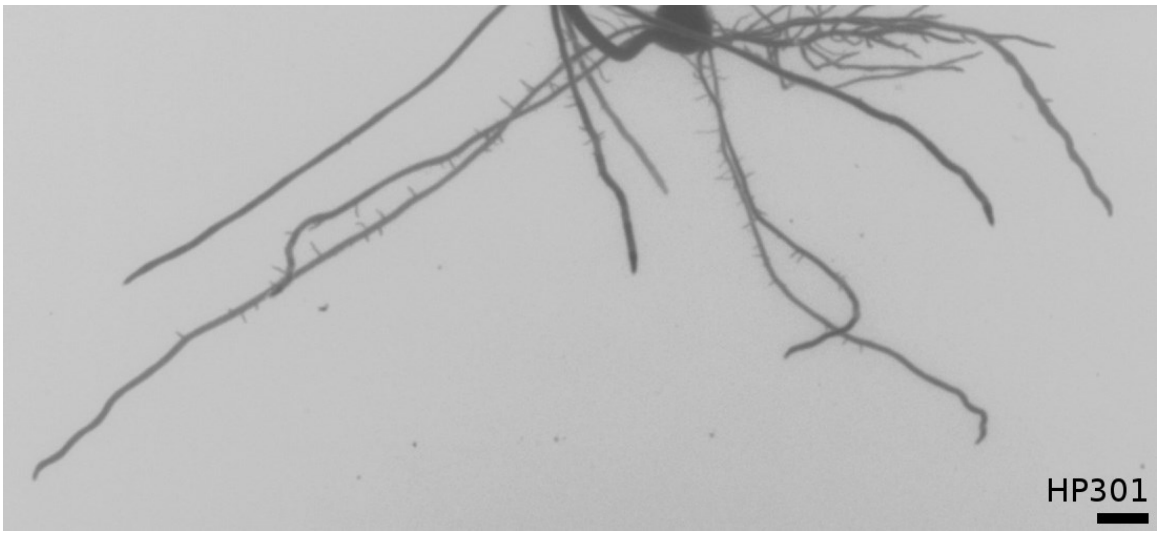

CML228

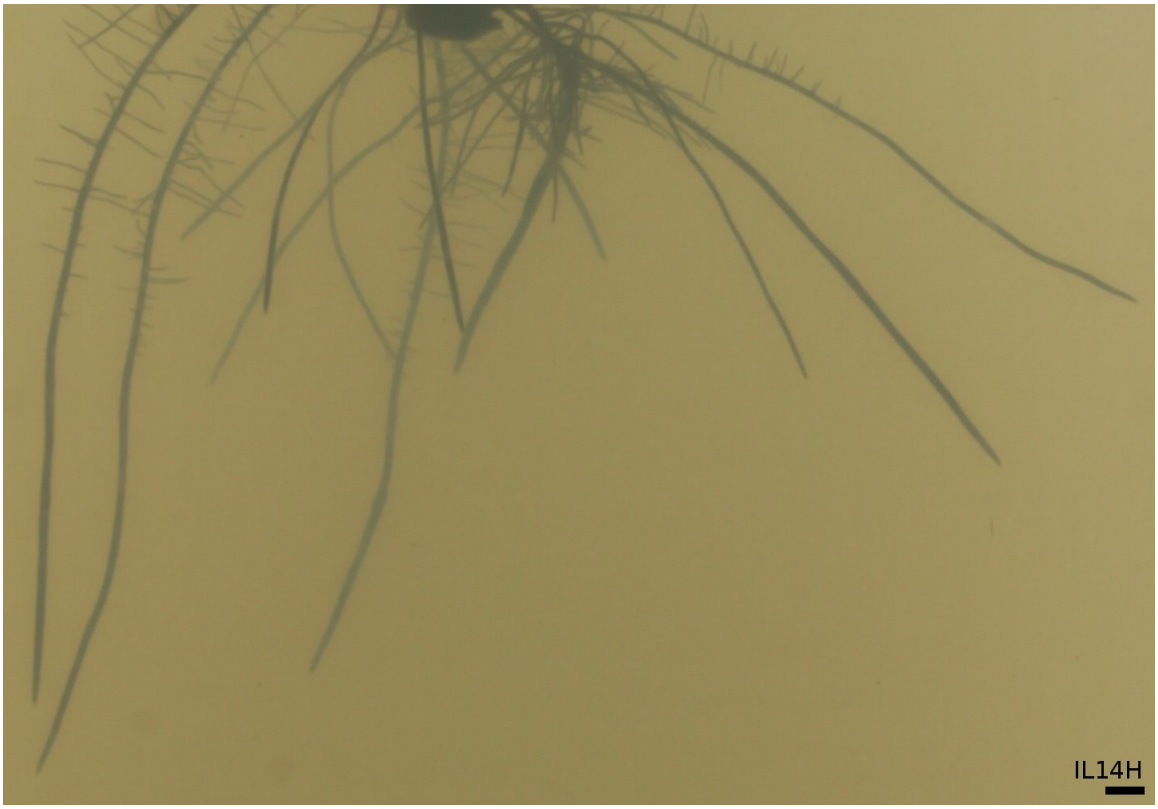

CML247

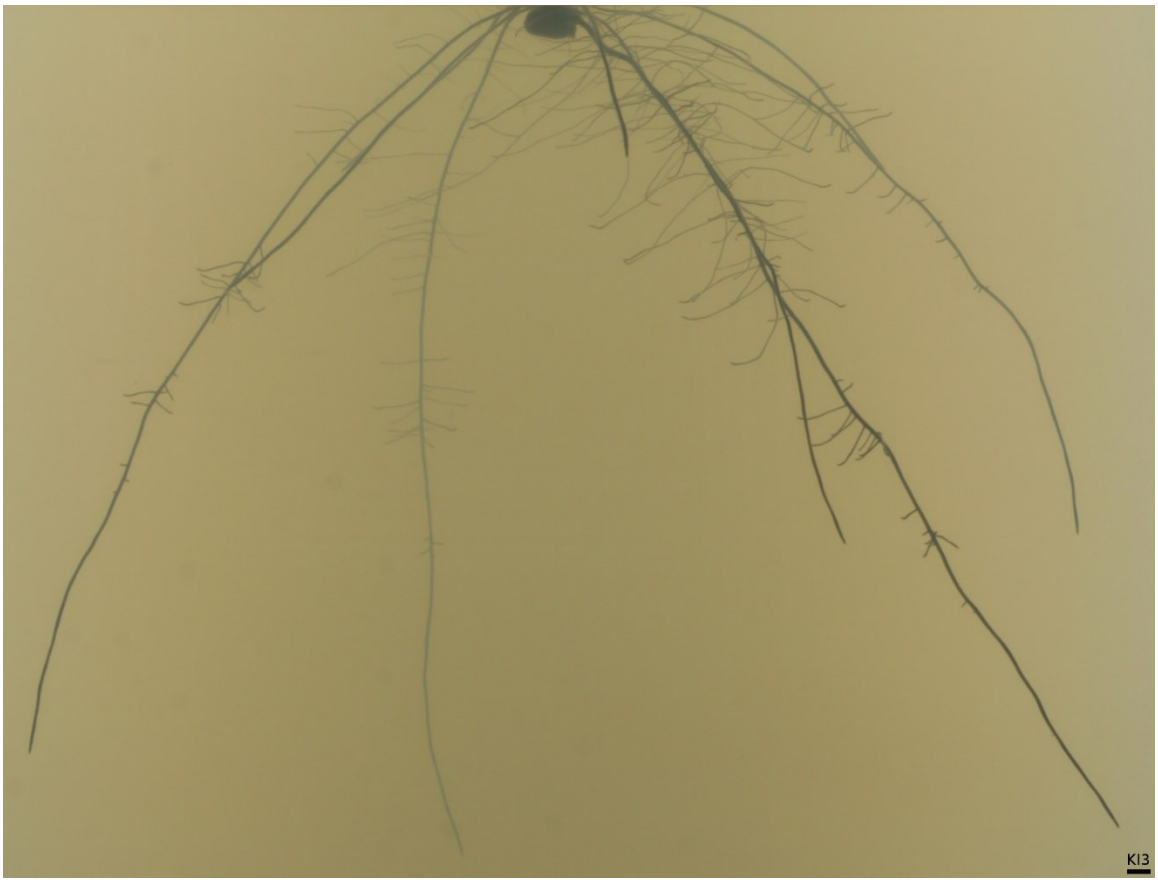

CML277

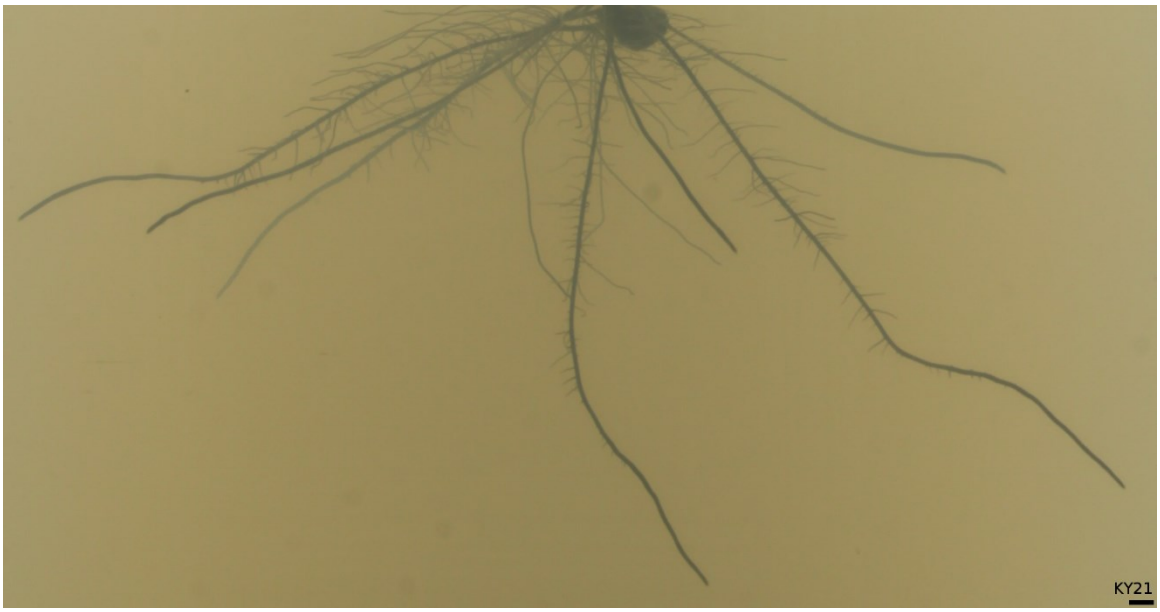

CML322

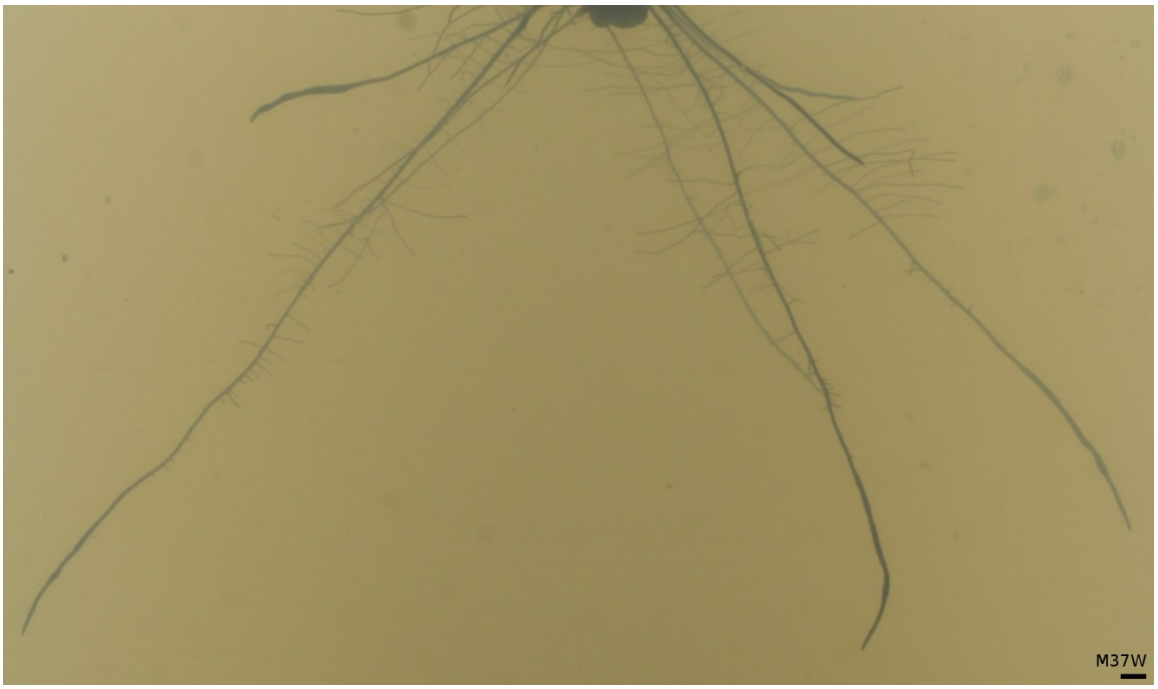

CML333

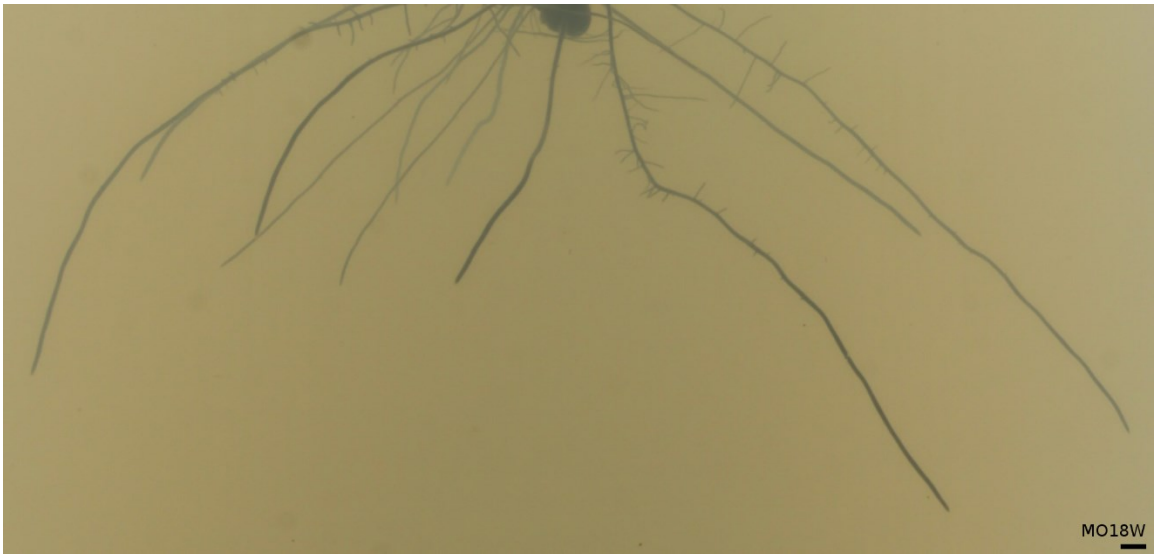

HP301

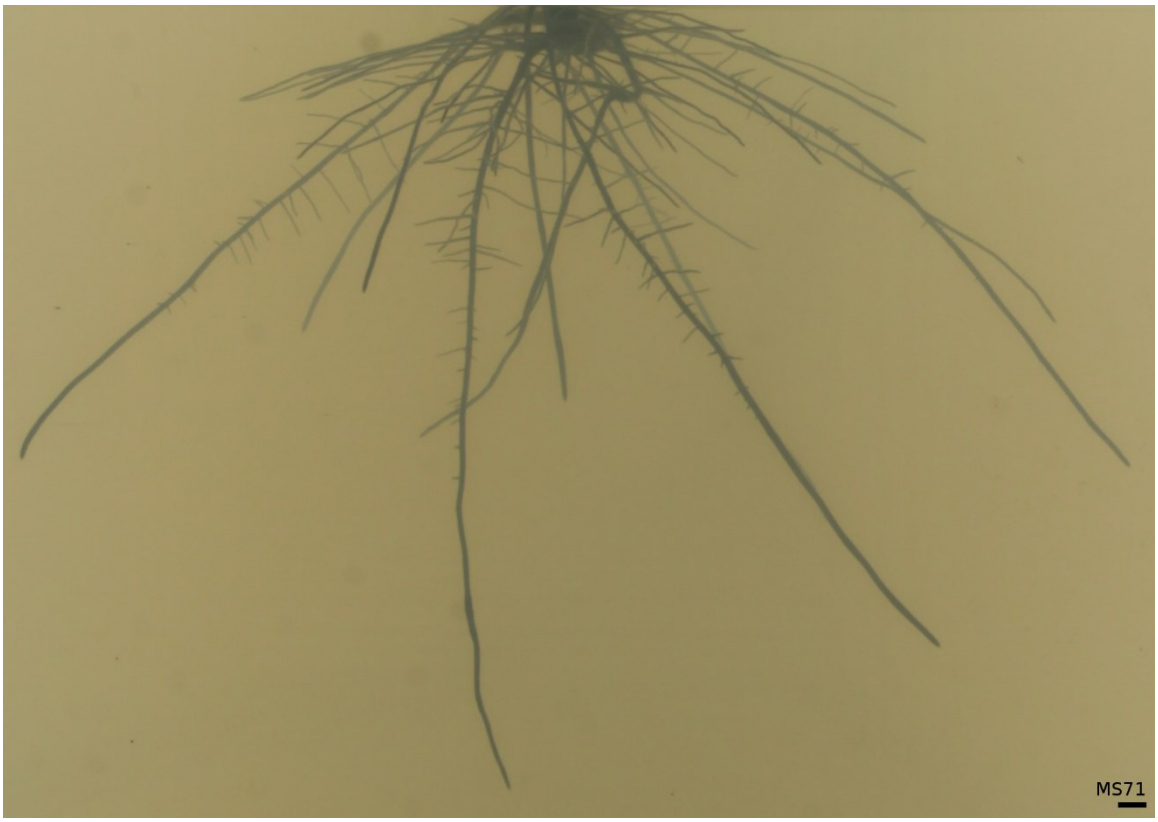

IL14H

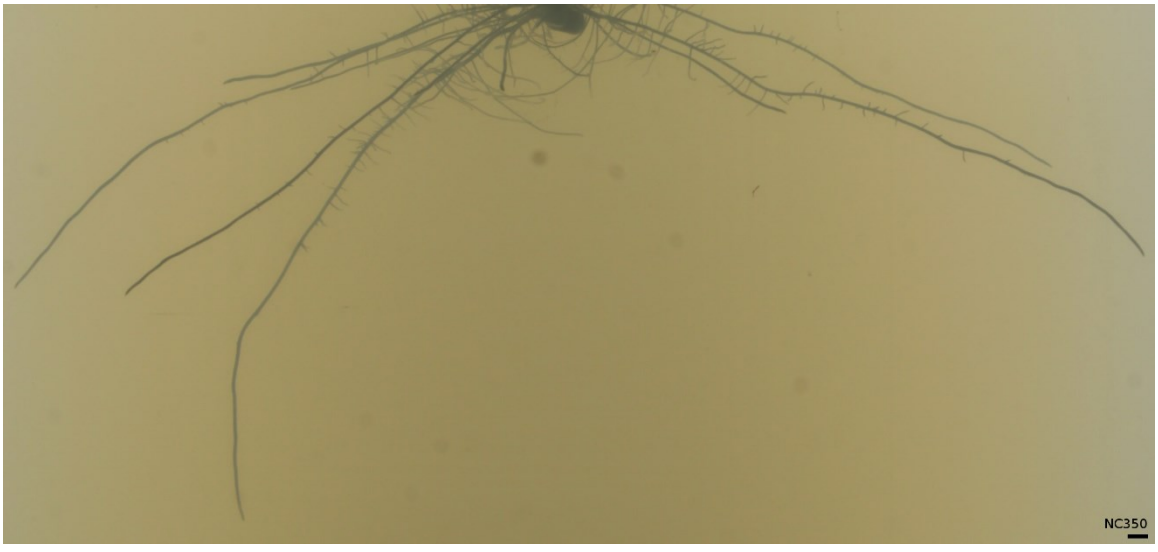

KI3


KY21

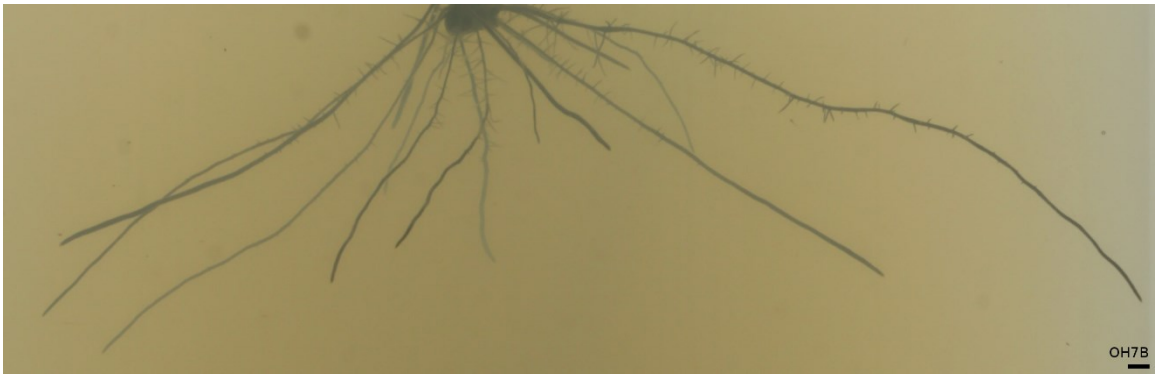

M37W

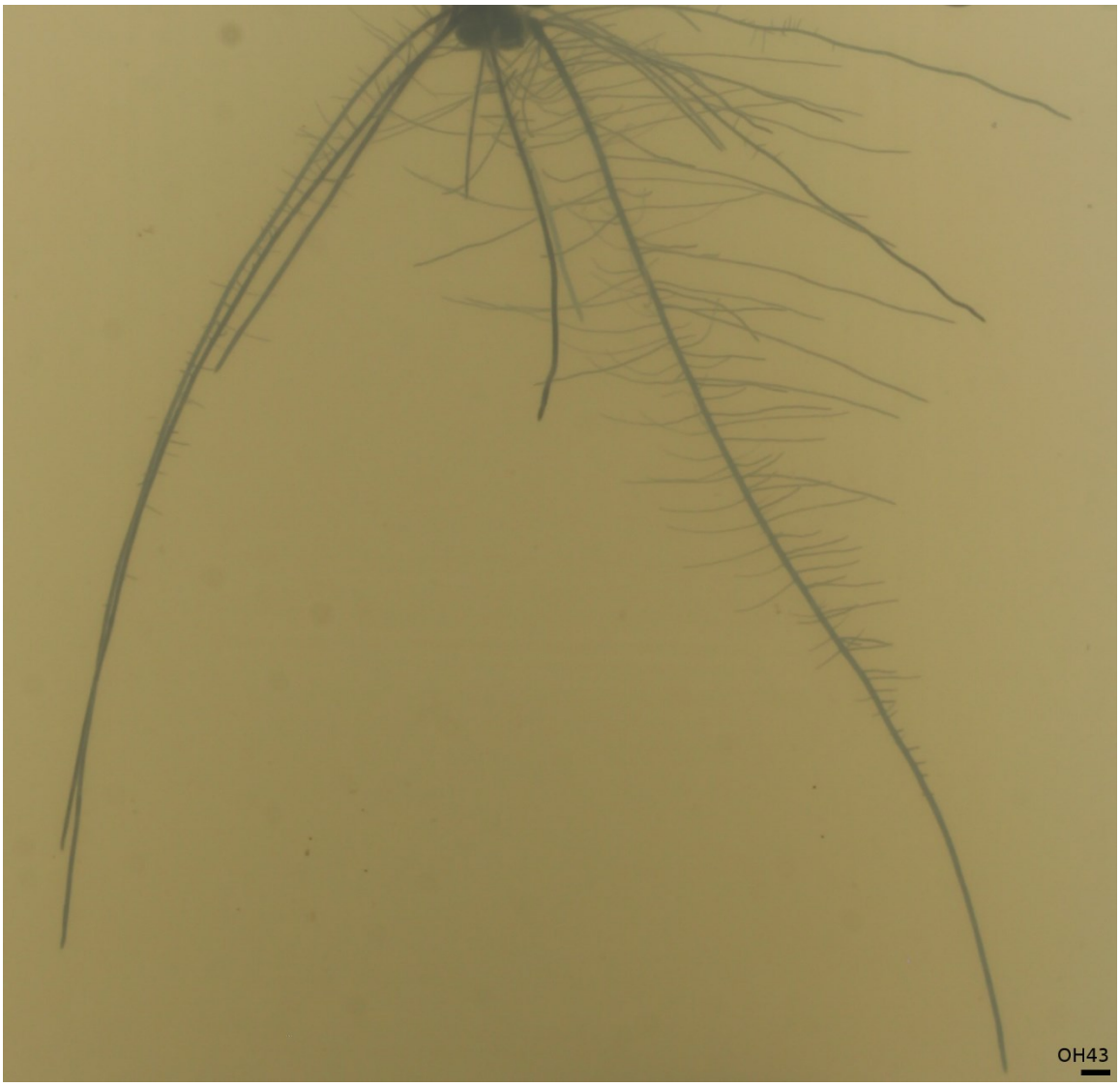

MO18W

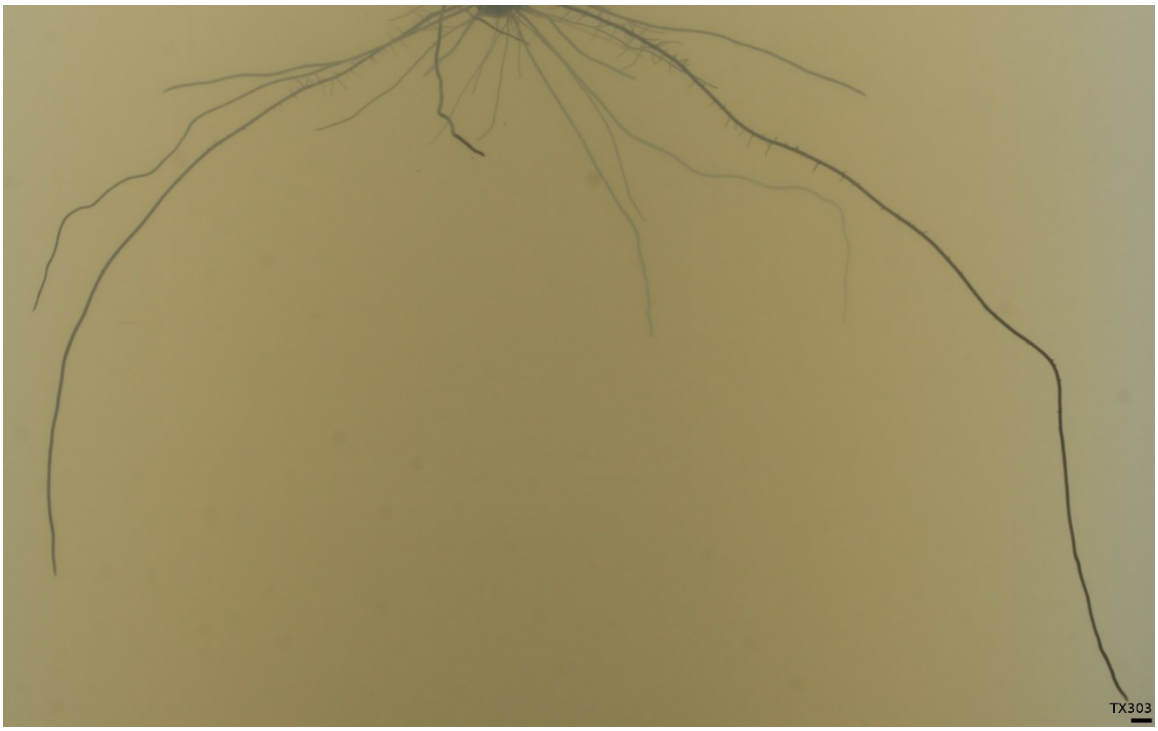

MS71

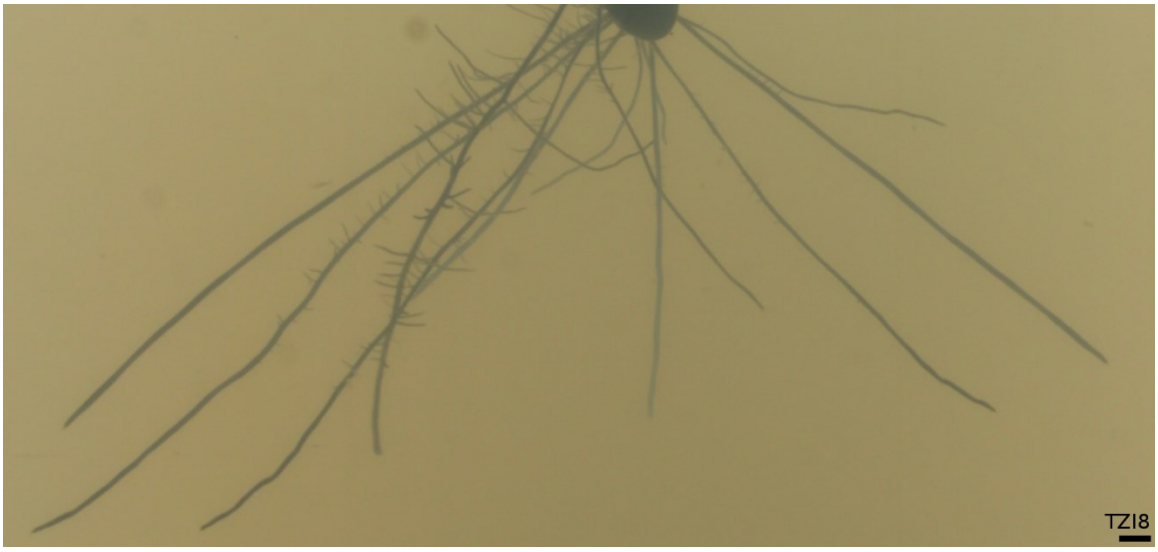

NC350


NC358


OH7B


OH43


TX303


TZI8

Appendix C. QTL list

TRAIT	CHR	MARKER	LR-SCORE	EFFECT SIZE	LEFT 2LOD	LEFT 1LOD	RIGHT 1LOD	RIGHT 2LOD	SIG LEVEL
depth(mm)_d08	10	56	53.282	0.238	50.521	55.543	58.532	60.374	0.01
minorellipseaxes(mm)_d06	10	54	47.670	0.198	46.776	46.961	54.805	55.157	0.01
minorellipseaxes(mm)_d08	10	60	47.756	0.191	50.873	55.662	62.330	62.816	0.01
depth(mm)_d06	10	56	49.215	0.189	50.689	55.554	61.112	62.154	0.01
networkconvexarea(mm^2)_d06	10	46	43.045	0.176	43.274	45.212	46.888	47.369	0.01
networkconvexarea(mm^2)_d08	10	46	44.429	0.170	43.077	43.775	47.150	47.463	0.01
majorellipseaxes(mm)_d06	10	44	39.146	0.156	41.981	43.308	46.592	47.739	0.01
lengthdistr(mm)_d06	6	12	38.571	0.154	7.144	9.507	15.321	15.855	0.01
totallength(mm)_d08	7	59	33.272	0.135	57.369	58.065	59.706	64.546	0.01
solidity_d06	2	122	30.953	0.127	113.416	116.814	122.802	130.250	0.01
widthdepthratio_d08	10	57	31.060	0.125	53.132	53.666	60.364	62.417	0.01
lengthdistr(mm)_d08	6	10	27.256	0.123	6.884	9.084	17.326	17.847	0.01
majorellipseaxes(mm)_d08	10	46	32.049	0.121	42.440	43.721	47.311	48.075	0.01
ellipseaxesaspectratio_d08	10	60	32.690	0.119	53.251	53.927	60.699	62.889	0.01
widthdepthratio_d06	10	57	30.047	0.116	53.603	55.424	60.056	61.165	0.01
totallength(mm)_d06	7	59	28.088	0.116	57.298	58.196	60.806	65.434	0.01
minorellipseaxes(mm)_d04	10	46	26.741	0.116	43.006	45.102	47.374	50.130	0.01
surfacearea(mm^2)_d04	10	46	26.441	0.114	42.756	45.319	47.242	47.801	0.01
maximumnumberofroots_d06	8	74	27.596	0.110	72.223	72.983	75.910	77.050	0.01
mediannumberofroots_d08	2	85	23.486	0.109	82.770	83.759	88.359	89.172	0.01
networkconvexarea(mm^2)_d06	2	121	28.235	0.109	112.246	115.568	122.227	122.711	0.01
depth(mm)_d04	10	46	26.427	0.108	43.341	45.330	51.270	56.062	0.01
ellipseaxesaspectratio_d06	10	58	24.890	0.107	50.609	55.519	64.160	66.246	0.01
majorellipseaxes(mm)_d06	2	117	28.604	0.107	105.470	113.930	122.030	127.319	0.01
ellipseaxesaspectratio_d08	7	117	30.295	0.106	111.584	112.816	122.467	130.548	0.01
networkarea(mm^2)_d04	10	46	25.230	0.106	44.389	45.347	47.193	47.672	0.01

perimeter(mm)_d06	7	59	25.552	0.106	57.647	58.399	60.173	62.907	0.01
solidity_d06	10	57	25.679	0.105	46.997	53.776	62.486	64.760	0.01
perimeter(mm)_d08	7	59	26.336	0.102	49.895	57.717	64.259	65.188	0.01
specificrootlength(1/mm^2)_d08	4	30	24.352	0.101	20.645	21.143	32.081	33.250	0.01
specificrootlength(1/mm^2)_d08	2	121	24.172	0.099	116.806	119.606	123.497	131.061	0.01
solidity_d08	10	58	24.386	0.098	46.813	50.315	66.105	66.334	0.01
specificrootlength(1/mm^2)_d06	8	13	23.144	0.098	2.622	11.519	14.492	15.595	0.01
averagerootwidth(mm)_d06	2	121	23.727	0.096	116.926	119.368	122.264	124.338	0.01
networkconvexarea(mm^2)_d04	10	46	24.342	0.096	43.131	45.299	47.158	47.579	0.01
majorellipseaxes(mm)_d04	2	121	23.843	0.095	109.944	111.486	122.480	123.327	0.01
solidity_d04	2	116	21.922	0.095	109.853	111.956	122.317	122.974	0.01
majorellipseaxes(mm)_d08	3	88	25.719	0.094	82.917	84.960	92.970	95.431	0.01
majorellipseaxes(mm)_d08	2	104	24.641	0.093	97.458	100.169	108.574	114.169	0.01
depth(mm)_d04	2	121	22.805	0.093	112.746	115.913	122.891	131.200	0.01
surfacearea(mm^2)_d06	10	46	21.770	0.091	42.118	44.886	47.226	47.811	0.01
majorellipseaxes(mm)_d04	10	46	22.505	0.091	43.677	45.470	47.397	51.383	0.01
networkconvexarea(mm^2)_d08	3	86	25.093	0.089	83.160	83.797	88.173	88.971	0.01
volume(mm^3)_d06	10	26	21.200	0.088	23.299	24.409	29.021	35.204	0.01
maxwidth(mm)_d06	3	88	20.770	0.087	78.844	84.843	92.570	95.524	0.01
averagerootwidth(mm)_d08	2	121	22.516	0.086	118.230	119.873	122.541	131.237	0.01
surfacearea(mm^2)_d08	7	59	21.329	0.086	49.273	50.110	65.044	65.822	0.01
depth(mm)_d08	2	128	22.277	0.085	119.563	127.293	130.029	130.923	0.01
surfacearea(mm^2)_d06	7	59	20.522	0.084	56.929	58.084	62.357	68.008	0.01
minorellipseaxes(mm)_d06	3	86	24.871	0.084	78.652	83.370	86.916	88.835	0.01
depth(mm)_d06	2	122	23.478	0.083	112.580	118.716	130.848	131.246	0.01
totallength(mm)_d04	9	65	19.864	0.082	56.959	61.734	67.748	70.760	0.01
networkarea(mm^2)_d06	10	46	19.682	0.082	41.940	42.990	47.253	47.791	0.01
averagerootwidth(mm)_d08	4	31	21.079	0.081	22.336	27.308	33.229	37.380	0.01
majorellipseaxes(mm)_d08	1	68	22.258	0.081	56.831	61.579	71.231	76.078	0.01

maxwidth(mm)_d04	10	46	17.956	0.080	35.781	39.062	47.217	49.064	0.05
networkarea(mm^2)_d08	7	59	19.734	0.079	49.249	50.015	64.681	65.654	0.01
averagerootwidth(mm)_d06	4	28	19.722	0.079	19.457	23.011	31.394	36.592	0.01
bushiness_d08	6	12	18.072	0.079	6.629	6.963	17.711	21.294	0.05
maxwidth(mm)_d06	10	46	18.757	0.079	35.666	39.900	47.082	47.764	0.05
perimeter(mm)_d08	4	30	20.863	0.079	19.907	21.714	31.549	36.414	0.01
networkarea(mm^2)_d08	10	26	19.127	0.078	22.653	23.574	27.369	30.720	0.05
maximumnumberofroots_d06	7	59	19.710	0.078	49.059	58.274	59.893	68.048	0.01
bushiness_d06	6	21	17.190	0.077	1.549	1.856	21.281	21.556	0.05
networkconvexarea(mm^2)_d04	2	116	20.241	0.077	109.976	112.094	122.191	122.792	0.05
networkconvexarea(mm^2)_d06	3	86	22.706	0.076	78.529	83.310	88.603	93.183	0.01
mediannumberofroots_d04	9	64	16.935	0.076	58.192	58.853	69.114	69.967	0.05
surfacearea(mm^2)_d08	10	26	18.606	0.075	22.230	23.484	27.483	30.792	0.05
perimeter(mm)_d04	9	65	18.153	0.075	60.528	61.814	68.925	70.674	0.05
networkarea(mm^2)_d06	3	86	18.665	0.075	78.769	83.682	89.321	93.722	0.05
averagerootwidth(mm)_d06	8	39	17.885	0.075	33.240	33.877	46.081	46.392	0.05
widthdepthratio_d06	7	117	20.006	0.075	109.970	112.646	124.175	131.014	0.01
bushiness_d04	1	0	16.940	0.075	0.000	0.000	3.589	6.799	0.05
widthdepthratio_d08	7	122	19.243	0.074	109.629	111.365	124.641	131.121	0.01
totallength(mm)_d08	4	27	19.686	0.074	21.143	21.964	31.541	36.178	0.01
depth(mm)_d06	1	201	21.275	0.074	198.360	198.898	202.218	208.000	0.01
lengthdistr(mm)_d06	1	191	15.830	0.074	182.788	185.765	199.259	200.392	0.05
volume(mm^3)_d06	9	37	17.792	0.073	30.831	31.644	39.637	46.239	0.05
widthdepthratio_d08	10	92	19.326	0.073	85.255	85.934	93.254	96.219	0.01
solidity_d08	1	76	18.261	0.072	65.493	65.820	77.459	85.468	0.05
widthdepthratio_d06	10	91	19.466	0.072	80.758	85.725	93.293	95.868	0.01
maximumnumberofroots_d08	4	28	18.874	0.072	21.965	24.706	29.969	35.166	0.05
majorellipseaxes(mm)_d06	1	64	19.874	0.071	53.811	59.908	70.671	73.024	0.01
majorellipseaxes(mm)_d06	3	88	19.759	0.071	78.136	83.292	92.812	96.208	0.01

minorellipseaxes(mm)_d06	2	121	20.722	0.071	109.985	112.643	122.696	131.249	0.01
solidity_d04	10	46	16.656	0.070	37.823	45.197	48.278	51.834	0.05
maximumnumberofroots_d08	2	83	18.377	0.070	78.710	81.662	89.449	92.534	0.05
averagerootwidth(mm)_d04	9	64	16.021	0.069	57.639	60.682	67.933	70.971	0.05
maximumnumberofroots_d08	7	59	18.087	0.069	49.528	56.575	62.060	66.256	0.05
networkarea(mm^2)_d06	7	60	16.733	0.067	48.966	57.745	65.750	68.288	0.05
maximumnumberofroots_d06	9	64	17.338	0.067	60.503	60.805	74.781	80.046	0.05
minorellipseaxes(mm)_d08	3	86	19.178	0.066	78.860	83.480	87.020	89.911	0.01
solidity_d08	3	96	17.126	0.066	85.636	95.040	99.285	101.040	0.05
networkconvexarea(mm^2)_d06	9	54	16.991	0.064	47.764	51.845	54.667	58.227	0.05
minorellipseaxes(mm)_d04	9	34	16.746	0.064	31.063	31.669	39.189	40.395	0.05
majorellipseaxes(mm)_d08	5	29	17.634	0.064	26.806	28.367	32.221	37.247	0.05
averagerootwidth(mm)_d08	8	38	16.239	0.063	33.134	33.780	39.890	46.361	0.05
depth(mm)_d04	7	107	15.545	0.062	96.418	105.858	108.850	117.296	0.05
totallength(mm)_d06	9	64	15.931	0.062	54.863	61.247	67.367	70.992	0.05
networkconvexarea(mm^2)_d08	9	54	17.863	0.062	48.821	51.905	54.792	55.877	0.05
totallength(mm)_d06	3	88	16.438	0.059	79.589	85.035	92.569	94.742	0.05
minorellipseaxes(mm)_d06	9	54	16.287	0.055	47.750	51.790	57.012	62.338	0.05

References

Abendroth, L.J., Elmore, R.W., Boyer, M.J., and Marlay, S.K. (2011). Corn Growth and Development.

Albanese, D., Visintainer, R., Merler, S., Riccadonna, S., Jurman, G., and Furlanello, C. (2012). mlpy: Machine Learning Python.

Armstrong, W., Bandle, R., and Jackson, M.B. (1994). Mechanisms of flood tolerance in plants. Acta Bot. Neerl. **43**: 307–358.

Bailey, D.W. (1971). Recombinant-inbred strains. An aid to finding identity, linkage, and function of histocompatibility and other genes. Transplantation **11**: 325–7.

Balint-Kurti, P.J. and Carson, M.L. (2006). Analysis of quantitative trait Loci for resistance to southern leaf blight in juvenile maize. Phytopathology **96**: 221–5.

Basten, C.J., Weir, B.S., and Zeng, Z.B. (2004). QTL Cartographer, Version 1.17 (Department of Statistics, North Carolina State University, Raleigh, NC).

Basten, C.J., Weir, B.S., and Zeng, Z.B. (1994). Zmap–a QTL cartographer. In Procceding of the 5th World Congress on Genetics Applied to Livestock Production: Computing Strategies and Software, pp. 65–66.

Buckler, E.S. et al. (2009). The genetic architecture of maize flowering time. Science **325**: 714–8.

Clark, R.T., MacCurdy, R.B., Jung, J.K., Shaff, J.E., McCouch, S.R., Aneshansley, D.J., and Kochian, L. V (2011). Three-dimensional root phenotyping with a novel imaging and software platform. Plant Physiol. **156**: 455–65.

Crow, J.F. (2007). Haldane, Bailey, Taylor and Recombinant-Inbred Lines. Genetics **176**: 729–732.

Doebley, J., Stec, A., Wendel, J., and Edwards, M. (1990). Genetic and morphological analysis of a maize-teosinte F2 population: implications for the origin of maize. Proc. Natl. Acad. Sci. U. S. A. **87**: 9888–92.

Dupuis, J. and Siegmund, D. (1999). Statistical methods for mapping quantitative trait loci from a dense set of markers. Genetics **151**: 373–86.

Duvick, D. (2005). Genetic progress in yield of United States maize (Zea mays L.). Maydica **50**: 193–202.

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J. (2008). LIBLINEAR: A Library for Large Linear Classification. J. Mach. Learn. Res.: 1871–1874.

Fang, S., Clark, R.T., Zheng, Y., Iyer-Pascuzzi, A.S., Weitz, J.S., Kochian, L. V, Edelsbrunner, H., Liao, H., and Benfey, P.N. (2013). Genotypic recognition and spatial responses by rice roots. Proc. Natl. Acad. Sci. U. S. A. **110**: 2670–5.

Ferrieri, A.P., Agtuca, B., Appel, H.M., Ferrieri, R.A., and Schultz, J.C. (2013). Temporal changes in allocation and partitioning of new carbon as (11)C elicited by simulated herbivory suggest that roots shape aboveground responses in Arabidopsis. Plant Physiol. **161**: 692–704.

Flint-Garcia, S.A., Thuillet, A.-C., Yu, J., Pressoir, G., Romero, S.M., Mitchell, S.E., Doebley, J., Kresovich, S., Goodman, M.M., and Buckler, E.S. (2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. **44**: 1054–64.

Fry, J.D. (1992). The Mixed-Model Analysis of Variance Applied to Quantitative Genetics: Biological Meaning of the Parameters. Evolution (N. Y). **46**: 540.

Galkovskyi, T. et al. (2012). GiA Roots: software for the high throughput analysis of plant root system architecture. BMC Plant Biol. **12**: 116.

Gamuyao, R., Chin, J.H., Pariasca-Tanaka, J., Pesaresi, P., Catausan, S., Dalid, C., Slamet-Loedin, I., Tecson-Mendoza, E.M., Wissuwa, M., and Heuer, S. (2012). The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. Nature **488**: 535–9.

Hammer, G.L., Dong, Z., McLean, G., Doherty, A., Messina, C., Schussler, J., Zinselmeier, C., Paszkiewicz, S., and Cooper, M. (2009). Can Changes in Canopy and/or Root System Architecture Explain Historical Maize Yield Trends in the U.S. Corn Belt? Crop Sci. **49**: 299.

Hansey, C.N., Vaillancourt, B., Sekhon, R.S., de Leon, N., Kaeppler, S.M., and Buell, C.R. (2012). Maize (Zea mays L.) genome diversity as revealed by RNA-sequencing. PLoS One **7**: e33071.

Hargreaves, C.E., Gregory, P.J., and Bengough, A.G. (2008). Measuring root traits in barley (Hordeum vulgare ssp. vulgare and ssp. spontaneum) seedlings using gel chambers, soil sacs and X-ray microtomography. Plant Soil **316**: 285–297.

Ho, M., Rosas, J., Brown, K., and Lynch, J.P. (2005). Root architectural tradeoffs for water and phosphorus acquisition. Funct. Plant Biol. **32**: 737–748.

Hoagland, D.R. and Arnon, D.I. (1950). The water-culture method for growing plants without soil. Calif. Agric. Exp. Stn. Circ.: 1–32.

Hochholdinger, F., Park, W.J., Sauer, M., and Woll, K. (2004). From weeds to crops: genetic analysis of root development in cereals. Trends Plant Sci. **9**: 42–8.

Holloway, B., Luck, S., Beatty, M., Rafalski, J.-A., and Li, B. (2011). Genome-wide expression quantitative trait loci (eQTL) analysis in maize. BMC Genomics **12**: 336.

Hund, A., Trachsel, S., and Stamp, P. (2009). Growth of axile and lateral roots of maize: I development of a phenotying platform. Plant Soil **325**: 335–349.

Iyer-Pascuzzi, A.S., Symonova, O., Mileyko, Y., Hao, Y., Belcher, H., Harer, J., Weitz, J.S., and Benfey, P.N. (2010). Imaging and analysis platform for automatic phenotyping and trait ranking of plant root systems. Plant Physiol. **152**: 1148–57.

Jahnke, S. et al. (2009). Combined MRI-PET dissects dynamic changes in plant structures and functions. Plant J. **59**: 634–44.

Jobbágy, E.G. and Jackson, R.B. (2001). The distribution of soil nutrients with depth: Global patterns and the imprint of plants. Biogeochemistry **53**: 51–77.

Lander, E.S. and Botstein, D. (1989). Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps. Genetics **121**: 185–199.

Lee, M., Sharopova, N., Beavis, W.D., Grant, D., Katt, M., Blair, D., and Hallauer, A. (2002). Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Mol. Biol. **48**: 453–61.

Li, Z., Gao, Q., Liu, Y., He, C., Zhang, X., and Zhang, J. (2011). Overexpression of transcription factor ZmPTF1 improves low phosphate tolerance of maize by regulating carbon metabolism and root growth. Planta **233**: 1129–43.

Liu, K., Goodman, M., Muse, S., Smith, J.S., Buckler, E., and Doebley, J. (2003). Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics **165**: 2117–28.

Lugtenberg, B. and Kamilova, F. (2009). Plant-growth-promoting rhizobacteria. Annu. Rev. Microbiol. **63**: 541–56.

Lynch, J.P. (1995). Root Architecture and Plant Productivity. Plant Physiol.: 7–13.

Lynch, J.P. (2013). Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems. Ann. Bot. **112**: 347–57.

Lyons, E., Pedersen, B., Kane, J., and Freeling, M. (2008). The Value of Nonmodel Genomes and an Example Using SynMap Within CoGe to Dissect the Hexaploidy that Predates the Rosids. Trop. Plant Biol. **1**: 181–190.

Magalhaes, J. V et al. (2007). A gene in the multidrug and toxic compound extrusion (MATE) family confers aluminum tolerance in sorghum. Nat. Genet. **39**: 1156–61.

Le Marié, C., Kirchgessner, N., Marschall, D., Walter, A., and Hund, A. (2014). Rhizoslides: paper-based growth system for non-destructive, high throughput phenotyping of root development by means of image analysis. Plant Methods **10**: 13.

Martin, B.A., Cerwick, S.F., and Reding, L.D. (1991). Physiology Basis for Inhibition of Maize Seed Germination by Flooding. Crop Sci. **31**: 1052.

Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez G, J., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. U. S. A. **99**: 6080–4.

Mauricio, R. (2001). Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. Nat. Rev. Genet. **2**: 370–81.

McMullen, M.D. et al. (2009). Genetic properties of the maize nested association mapping population. Science **325**: 737–40.

Pearson, T.A. and Manolio, T.A. (2008). How to interpret a genome-wide association study. JAMA **299**: 1335–44.

Peng, J. et al. (1999). "Green revolution" genes encode mutant gibberellin response modulators. Nature **400**: 256–261.

Reiter, R.S., Williams, J.G., Feldmann, K.A., Rafalski, J.A., Tingey, S. V, and Scolnik, P.A. (1992). Global and local genome mapping in Arabidopsis thaliana by using recombinant inbred lines and random amplified polymorphic DNAs. Proc. Natl. Acad. Sci. U. S. A. **89**: 1477–81.

Rick, C.M. and Smith, P.G. (1953). Novel Variation in Tomato Species Hybrids. Am. Nat.: 359–373.

SAS Institue Inc (2014). JMP®, Version 11.

Sasaki, A., Ashikari, M., Ueguchi-Tanaka, M., Itoh, H., Nishimura, A., Swapan, D., Ishiyama, K., Saito, T., Kobayashi, M., Khush, G.S., Kitano, H., and Matsuoka, M. (2002). Green revolution: a mutant gibberellin-synthesis gene in rice. Nature **416**: 701–2.

Schnable, P.S. et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. Science **326**: 1112–5.

Shull, G. (1908). The composition of a field of maize. Rep Am Breeders Assoc: 296–301.

Smith, C.W., Betran, J., and Runge, E.C.A. (2004). Corn: Origin, History, Technology, and Production (John Wiley & Sons).

Suslow, T. V. and Schroth, M.N. (1982). Role of Deleterious Rhizobacteria as Minor Pathogens in Reducing Crop Growth. Phytopathology **72**: 111.

Topp, C.N. et al. (2013). 3D phenotyping and quantitative trait locus mapping identify core regions of the rice genome controlling root architecture. Proc. Natl. Acad. Sci. U. S. A. **110**: E1695–704.

Trachsel, S., Kaeppler, S.M., Brown, K.M., and Lynch, J.P. (2010). Shovelomics: high throughput phenotyping of maize (Zea mays L.) root architecture in the field. Plant Soil **341**: 75–87.

Tuberosa, R., Salvi, S., Sanguineti, M.C., Maccaferri, M., Giuliani, S., and Landi, P. (2003). Searching for quantitative trait loci controlling root traits in maize: a critical appraisal. Plant Soil **255**: 35–54.

Tuberosa, R., Sanguineti, M.C., Landi, P., Giuliani, M.M., Salvi, S., and Conti, S. (2002). Identification of QTLs for root characteristics in maize grown in hydroponics and analysis of their overlap with QTLs for grain yield in the field at two water regimes. Plant Mol. Biol. **48**: 697–712.

Tuinstra, M.R., Ejeta, G., and Goldsbrough, P.B. (1997). Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. TAG Theor. Appl. Genet. **95**: 1005–1011.

Udomprasert, N., Kijjanon, J., Thiraporn, R., and Machuay, A. (1997). Effects of water deficit at tasseling on proline and ABA levels and yield of corn.

Uga, Y. et al. (2013). Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. Nat. Genet. **45**: 1097–102.

Visscher, P.M., Thompson, R., and Haley, C.S. (1996). Confidence Intervals in QTL Mapping by Bootstrapping. Genetics **143**: 1013–1020.

Wang, H., Nussbaum-Wagler, T., Li, B., Zhao, Q., Vigouroux, Y., Faller, M., Bomblies, K., Lukens, L., and Doebley, J.F. (2005). The origin of the naked grains of maize. Nature **436**: 714–9.

Weaver, J.E. and Bruner, W.E. (1927). Root Development of Vegetable Crops.

Weller, J.I. and Soller, M. (2004). An analytical formula to estimate confidence interval of QTL location with a saturated genetic map as a function of experimental design. Theor. Appl. Genet. **109**: 1224–9.

Wright, S.I., Bi, I.V., Schroeder, S.G., Yamasaki, M., Doebley, J.F., McMullen, M.D., and Gaut, B.S. (2005). The effects of artificial selection on the maize genome. Science **308**: 1310–4.

Yoshida, S., Forno, D., J, Cock, J., and Ka, G. (1976). Laboratory Manual for Physiological Studies of Rice 3rd ed. (International Rice Research Institute).

Yu, J., Holland, J.B., McMullen, M.D., and Buckler, E.S. (2008). Genetic design and statistical power of nested association mapping in maize. Genetics **178**: 539–51.

Zeng, Z.B. (1994). Precision Mapping of Quantitative Trait Loci. Genetics **136**: 1457–1468.

Zeng, Z.B. (1993). Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. U. S. A. **90**: 10972–6.

Zheng, Y., Gu, S., Edelsbrunner, H., Tomasi, C., and Benfey, P.N. (2011). Detailed reconstruction of 3D plant root shape. In International Conference on Computer Vision, pp. 1–8.

Biography

Paul Roman Zurek was born in a Rylowa, Poland on March 30, 1982. He grew up on a small farm in Poland and relocated with his whole family to Chicago when he was 13. He graduated from the St. Patrick High School in 2002. Paul received his Bachelor of Science in Biological Science with a minor in Chemistry from Loyola University is 2006. He received his Master of Science in Plant Genetics in the lab of Dr. F. Bryan Pickett from Loyola University in 2008.